

## O'Bryen, Barbara

---

**From:** Swope, Sheridan  
**Sent:** Monday, July 14, 2003 2:07 PM  
**To:** O'Bryen, Barbara  
**Subject:** FW: 09966880

**Importance:** High

Barb, May I get a copy of the search results for 09966880?

THANKS!

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1652  
Recombinant Enzymes  
sheridan.swope@uspto.gov  
703-305-1696 (voice)  
703-308-3014 (FAX)  
Mailbox: CM1 Rm10D01  
Office: CM1 Rm12D12

-----Original Message-----

**From:** STIC-Biotech/ChemLib  
**Sent:** Monday, July 14, 2003 1:45 PM  
**To:** Swope, Sheridan  
**Subject:** RE: 09966880

Please check your mail box or call Barb O'Bryen. She completed the search on 7/8/03

-----Original Message-----

**From:** Swope, Sheridan  
**Sent:** Monday, July 14, 2003 1:38 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** FW: 09966880

WHAT IS THE STATUS OF THIS RUSH REQUEST?

-----Original Message-----

**From:** Swope, Sheridan  
**Sent:** Friday, July 04, 2003 2:26 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** FW: 09966880

-----Original Message-----

**From:** Richter, Johann  
**Sent:** Friday, July 04, 2003 2:06 PM  
**To:** Swope, Sheridan; STIC-Biotech/ChemLib  
**Cc:** Chan, Christina  
**Subject:** RE: 09966880

Approved.

*Johann R. Richter, Ph.D., Esq.  
Supervisory Patent Examiner  
Biotechnology and Organic Chemistry  
Art Unit 1621  
703-308-4532*

-----Original Message-----

**From:** Swope, Sheridan  
**Sent:** Friday, July 04, 2003 1:59 PM

**To:** STIC-Biotech/ChemLib  
**Cc:** Chan, Christina; Richter, Johann  
**Subject:** 09966880  
**Importance:** High

I made a mistake on the original request for this search.  
May I have this rushed?

For 09966880, pls search:

SID 7 oligo against the NT and AA data bases.

SID 8 oligo against the NT and AA data bases.

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1652  
Recombinant Enzymes  
sheridan.swope@uspto.gov  
703-305-1696 (voice)  
703-308-3014 (FAX)  
Mailbox: CM1 Rm10D01  
Office: CM1 Rm12D12

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: July 7, 2003, 20:38:32 ; Search time 73.5 Seconds  
(without alignments)  
10217.694 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 897  
Sequence: 1 agagacacatcattatga.....aaaaaaaaaaaaaaaa 2818

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus.n2p.model -DEV=rlp  
-Q=/cgn2\_1/USPTO.spool/US09966880/runatc\_07072003\_142232\_22058/app\_query.fasta\_1.3015  
-DB=A.geneseq.101002 -QFMT=fastan -SUFFIX=01g.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -STRAT=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880.eccn\_1\_1.100.&runatc\_07072003\_142232\_22058 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEBUFFER -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

A.Geneseq.101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	198	22.1	198	21	Human activation-1
2	50	5.6	50	22	Human immune/haema
3	42	4.7	198	21	Mouse activation-1
4	24	2.7	75	22	Human polypeptide
5	23	2.6	68	22	Human secreted pro
6	23	2.6	88	22	Human secreted pro
7	23	2.6	95	22	Human secreted pro
8	23	2.6	102	22	Human polypeptide
9	23	2.6	103	21	Human secreted pro
10	23	2.6	107	22	Human polypeptide
11	23	2.6	117	22	Human polypeptide
12	23	2.6	211	22	Human polypeptide
13	22	2.4	264	22	Novel human diagno
14	22	2.4	264	22	Novel human diagno
15	21	2.3	32	22	Novel human secret
16	21	2.3	32	22	Novel human secret
17	21	2.3	177	22	Human ribosomal pr
18	20	2.2	40	22	Human polypeptide
19	20	2.2	41	22	Human polypeptide
20	20	2.2	70	21	Human secreted pro
21	20	2.2	96	22	Human polypeptide
22	19	2.1	38	22	Novel human secret
23	19	2.1	44	23	Human ovarian anti
24	19	2.1	54	21	Human secreted pro
25	19	2.1	59	22	Human colon cancer
26	19	2.1	85	22	Human polypeptide
27	19	2.1	88	21	Human secreted pro
28	19	2.1	90	22	Human protein sequ
29	19	2.1	90	22	Human metabolic 3-
30	19	2.1	107	22	Novel human secret
31	19	2.1	108	22	Human secreted pro
32	19	2.1	113	22	Human polypeptide
33	19	2.1	121	22	Novel human secret
34	19	2.1	128	22	Human polypeptide
35	18	2.0	36	22	Human polypeptide
36	18	2.0	44	21	Lung cancer associ
37	18	2.0	44	22	Human polypeptide
38	18	2.0	45	21	Gene 7 human secret
39	18	2.0	65	21	Gene 43 human secret
40	18	2.0	65	21	Human secreted pro
41	18	2.0	66	21	Human secreted pro
42	18	2.0	67	21	Gene 17 human secr
43	18	2.0	71	22	Gene 4 human secre
44	18	2.0	76	21	Human polypeptide
45	18	2.0	79	21	Human secreted pro

## ALIGNMENTS

RESULT 1  
AAB24198  
ID AAB24198 standard: Protein; 198 AA.  
XX AAB24198;  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase SEQ ID NO:8.  
XX  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antineutemic; antiastrumatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW Iga nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;

KM auto immunodeficiency syndrome; IgG subclass selection disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
FE 28-MAR-2000; 2000WO-JP01918.  
XX  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178939.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
PA (NIBS ) JAPAN TOBACCO INC.  
PA (HONJ/) HONJO T.  
PI Honjo T, Muramatsu M;  
XX  
XX WPI; 2000-611715/58.  
DR N-PSDB; AAC55312.  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
PS Claim 1; Page 140-141; 174pp; Japanese.  
XX  
XX The present sequence is human activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,  
CC antianemic, antisthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency,  
CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DisGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders.  
XX  
SQ Sequence 198 AA:  
  
Alignment Scores:  
Pred. No.: 6.59e-193 Length: 198  
Score: 198.00 Matches: 198  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 22.07% Indels: 0  
Gaps: 0  
  
US-09-966-880A-7 (1-2818) x AAB24198 (1-198)  
QY 80 ATGACAGCCTCTTGATGAACCGAGAGATTTCTTACCAATTCATAAATGTCGCTCG 139  
DB 1 Metaspserleuenuetnsaargylvspheleryglphelyasnlvalargtpr 20  
QY 140 GCTAAGGCTGGCGTGAAGCTGCTGCTAGCTAGTGAAGAGGGCTGACAGTGCTACA 199  
DB 21 Alalysgllyargarglunthryrleucystyrvalvallyalargargysperalathr 40  
QY 200 TCCCTTCACTGGACTTGGTTATCTTCGCAATAGACAGGCTGCCACGTGAATGCTC 259  
DB 41 Serpheserleuasphegilyrleuargasnlyasnglycyshtslvalgileuenu 60  
QY 260 TTCTTCGCTACATCTCGGACTGGGACTAGACCCCTGGCCGCTGCTACCGCTGACTG 319  
DB 61 Pheleuargtyrilleseraspripaspseuaspptrogllyargcystyrargvalthartpr 80

QY 320 TTCACCTCTCTGAGAGCCCTGCTAGACTGTGCCGACATGTGGCCGACTTCTGCGAGG 379  
DB 81 Phertherserprserprocystryaspoxalarghlsvalalaspheleuarggly 100  
QY 380 AACCCCAACCTCATCTGAGAGATTTTCACGCGCGCTCTACTCTGTGAGGACCCAG 439  
DB 101 AsnproasnleuaserleuargllepheThAlaargleuYrPheCysgluasparglys 120  
QY 440 GCTGAGCCCGAGGGGCTGGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 499  
DB 121 AlagluProgluglyleuargargleuHlsarglacllyalgnllealilemetThr 140  
QY 500 TTCAAAGATTATTTTCTACTGCTGGAATCTTTTBAAAAACATGAAACATTTCCAA 559  
DB 141 PheylsasprlyrpheryrcysatrpasnthrPhevalgluasnhsiglunargthrPheLys 160  
QY 560 GCGTGGGAAGGGCGTCATGAATTCAGTTCGCTCTCCAGACAGCTTCGGCCATCTT 619  
DB 161 AlatrpgluglyleuHlsiglunserValalargleuserarglnleuarglileu 180  
QY 620 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTTGTTGGACTT 673  
DB 181 LeuProleuTyrgluValaspaspseuargspalapheargthrleuGllyeu 198  
  
RESULT 2  
ID AAM89038 standard; Protein; 50 AA.  
XX  
AC AAM89038;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen SEQ ID NO:16631.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.



	PR	14-AUG-2000;	2000US-0225758.
	PR	14-AUG-2000;	2000US-0225759.
	PR	18-AUG-2000;	2000US-0226779.
	PR	22-AUG-2000;	2000US-0226861.
	PR	22-AUG-2000;	2000US-0227182.
	PR	23-AUG-2000;	2000US-0227009.
	PR	30-AUG-2000;	2000US-0228924.
	PR	01-SEP-2000;	2000US-0229287.
	PR	01-SEP-2000;	2000US-0229343.
	PR	01-SEP-2000;	2000US-0229344.
	PR	01-SEP-2000;	2000US-0229345.
	PR	05-SEP-2000;	2000US-0229509.
	PR	05-SEP-2000;	2000US-0229513.
	PR	06-SEP-2000;	2000US-0230437.
	PR	06-SEP-2000;	2000US-0230438.
	PR	08-SEP-2000;	2000US-0231242.
	PR	08-SEP-2000;	2000US-0231243.
	PR	08-SEP-2000;	2000US-0231244.
	PR	08-SEP-2000;	2000US-0231413.
	PR	08-SEP-2000;	2000US-0231414.
	PR	08-SEP-2000;	2000US-0232080.
	PR	12-SEP-2000;	2000US-0232081.
	PR	12-SEP-2000;	2000US-0231968.
	PR	14-SEP-2000;	2000US-0232397.
	PR	14-SEP-2000;	2000US-0232398.
	PR	14-SEP-2000;	2000US-0232399.
	PR	14-SEP-2000;	2000US-0232401.
	PR	14-SEP-2000;	2000US-0232401.
	PR	14-SEP-2000;	2000US-0233063.
	PR	14-SEP-2000;	2000US-0233064.
	PR	14-SEP-2000;	2000US-0233065.
	PR	21-SEP-2000;	2000US-0234223.
	PR	21-SEP-2000;	2000US-0234274.
	PR	25-SEP-2000;	2000US-0234997.
	PR	25-SEP-2000;	2000US-0234998.
	PR	26-SEP-2000;	2000US-0235464.
	PR	27-SEP-2000;	2000US-0235834.
	PR	27-SEP-2000;	2000US-0235836.
	PR	29-SEP-2000;	2000US-0236337.
	PR	29-SEP-2000;	2000US-0236367.
	PR	29-SEP-2000;	2000US-0236368.
	PR	29-SEP-2000;	2000US-0236369.
	PR	29-SEP-2000;	2000US-0236370.
	PR	02-OCT-2000;	2000US-0236802.
	PR	02-OCT-2000;	2000US-0237037.
	PR	02-OCT-2000;	2000US-0237038.
	PR	02-OCT-2000;	2000US-0237039.
	PR	02-OCT-2000;	2000US-0237040.
	PR	13-OCT-2000;	2000US-0239393.
	PR	13-OCT-2000;	2000US-0239395.
	PR	20-OCT-2000;	2000US-0240960.
	PR	20-OCT-2000;	2000US-0241221.
	PR	20-OCT-2000;	2000US-0241785.
	PR	20-OCT-2000;	2000US-0241786.
	PR	20-OCT-2000;	2000US-0241787.
	PR	20-OCT-2000;	2000US-0241808.
	PR	20-OCT-2000;	2000US-0241809.
	PR	20-OCT-2000;	2000US-0241826.
	PR	01-NOV-2000;	2000US-0246417.
	PR	01-NOV-2000;	2000US-0246474.
	PR	08-NOV-2000;	2000US-0246475.
	PR	08-NOV-2000;	2000US-0246476.
	PR	08-NOV-2000;	2000US-0246477.
	PR	08-NOV-2000;	2000US-0246478.
	PR	08-NOV-2000;	2000US-0246523.
	PR	08-NOV-2000;	2000US-0246524.
	PR	08-NOV-2000;	2000US-0246525.
	PR	08-NOV-2000;	2000US-0246526.
	PR	08-NOV-2000;	2000US-0246527.
	PR	08-NOV-2000;	2000US-0246528.
	PR	08-NOV-2000;	2000US-0246532.
	PR	08-NOV-2000;	2000US-0246609.
	PR	14-AUG-2000;	2000US-0225758.
	PR	14-AUG-2000;	2000US-0225759.
	PR	18-AUG-2000;	2000US-0226779.
	PR	22-AUG-2000;	2000US-0226861.
	PR	22-AUG-2000;	2000US-0227182.
	PR	23-AUG-2000;	2000US-0227009.
	PR	30-AUG-2000;	2000US-0228924.
	PR	01-SEP-2000;	2000US-0229287.
	PR	01-SEP-2000;	2000US-0229343.
	PR	01-SEP-2000;	2000US-0229344.
	PR	01-SEP-2000;	2000US-0229345.
	PR	05-SEP-2000;	2000US-0229509.
	PR	05-SEP-2000;	2000US-0229513.
	PR	06-SEP-2000;	2000US-0230437.
	PR	06-SEP-2000;	2000US-0230438.
	PR	08-SEP-2000;	2000US-0231242.
	PR	08-SEP-2000;	2

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.57% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-966-880a-7 (1-2818) x AAM89038 (1-50)  
 QY 1470 AAACACCACCAACTTCACATATATAGCAACAATGGAGAGAGTGTGAT 1529  
 Db 1 LysHisProclnhrserHisIleIleSerLysGlnLeuGlnLysGlySerLeuasn 20  
 QY 1530 GTTGGGAGAGAGAAAATCATTTGGCTCTGCTGCTTCATCTGAAATGCCAATCA 1589  
 Db 21 ValGlyGlnArgLysIleLysTrpLeuSerTrpValSerSerSerGlnLysCysGlnser 40  
 QY 1590 GGTCAAGTTTGTTCATATTTGTATGTG 1619  
 Db 41 GlyGlnGlyLeuLeuHisPheValCysVal 50  
 RESULT 3  
 AAB24197  
 ID AAB24197 standard; Protein: 198 AA.  
 AC AAB24197;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Mouse activation-induced cytidine deaminase SEQ ID NO:2.  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; anti-allergic;  
 KW antiaiemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; Disgeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder.  
 OS Mus musculus.  
 XX  
 OS WO200058480-A1.  
 PN  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP01918.  
 XX  
 PR 29-MAR-1999; 99JP-0087192.  
 XX  
 PR 24-JUN-1999; 99JP-0178989.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX  
 PA (NIBS) JAPAN TOBACCO INC.  
 PA (HONT/) HONTO T.  
 PI  
 XX  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI: 2000-611715/58.  
 DR N-PSDB: AAC55307.  
 XX  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 PT  
 XX  
 XX Claim 1; Page 131-132; 174pp; Japanese.  
 PS  
 XX The present sequence is mouse activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antiaiemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IGA) deficiency

CC disease, IGA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, Disgeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC Ige disorder, and Igg subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX  
 SQ Sequence 198 AA;  
 XX  
 Alignment Scores:  
 Pred. No.: 1.77e-33 Length: 198  
 Score: 42.00 Matches: 42  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.68% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-966-880a-7 (1-2818) x AAB24197 (1-198)  
 QY 239 GCGTCCACAGTGAATTGCTCTCTCCGCTACATCTCGAGCTGGAGACCTGCG 298  
 Db 54 GlyCysHisValGlnLeuLeuPheLeuArgTrpIleSerAspTrpAspLeuAspProGly 73  
 QY 299 CGCTGCTACCGGCGTACCTGCTTCACTCTCGAGCCCTGCTACGACTGCCCCGACAT 358  
 Db 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93  
 QY 359 GTGGCC 364  
 Db 94 ValAla 95  
 RESULT 4  
 AA007022  
 ID AA007022 standard; Protein: 75 AA.  
 AC AA007022;  
 XX  
 AC AA007022;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide seq ID NO 20914.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 PN  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PA  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-514838/56.  
 DR N-PSDB: AAI86953.  
 XX  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 PT  
 XX  
 PS Claim 20; SEQ ID NO 20914; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and

CC the encoded proteins (AA00010-AA01310) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polypeptides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 75 AA:  
  
Alignment Scores:  
Pred. No.: 4.62e-15 Length: 75  
Score: 24.00 Matches: 24  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.67% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-966-880a-7 (1-2818) x AA007022 (1-75)  
  
QY 2054 CTGCACCTCCACCTCCTGGGTTCAAGCATTCTCTGCTCAGCTCCAGCTCCAGTACTCG 1995  
Db 17 LeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrp 36  
  
QY 1994 GATTACAGGTGC 1983  
Db 37 AspTyrArgCys 40  
  
RESULT 5  
ID AAB64608 standard; Protein: 68 AA.  
XX  
AC AAB64608;  
XX  
DT 22-MAR-2001 (first entry)  
XX  
DE Human secreted protein BLAST search protein SEQ ID NO: 118.  
XX  
DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW caduatic; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX  
OS Homo sapiens.  
XX  
PN WO200077197-A1.  
XX  
PD 21-DEC-2000.  
XX  
PF 01-JUN-2000; 2000WO-US14934.  
XX  
PR 11-JUN-1999; 9905-0138599.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PA (ROSE/) ROSEN C A.  
XX  
PI Rosen CA, Ruben SM, Komatsoulis GA;  
XX  
DR WPI; 2001-032312/04.  
XX  
PT Isolated nucleic acid molecule encoding a human secreted protein is  
XX used in preventing, treating or ameliorating a medical condition -  
XX  
PS Disclosure; Page 513; 558pp; English.  
XX  
XX The invention relates to the isolation of genes AAF32757-F32803 encoding

CC the human secreted proteins AAB64549-B64594. The sequence is used as a  
CC query sequence for doing BLASTX searches to identify homologous  
CC sequences. The genes and proteins are useful for preventing,  
CC ameliorating or treating medical conditions, e.g. by protein or gene  
CC therapy. The genes are isolated from a range of human tissues disclosed  
CC in the specification. The nucleic acids, proteins, antibodies and  
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal  
CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's  
CC disease, allergies, autoimmune haemolytic anaemia, autoimmune  
CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,  
CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular  
CC disorders such as myocardial ischaemias; (d) wound healing; (e)  
CC neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections.  
XX  
SQ Sequence 68 AA:  
  
Alignment Scores:  
Pred. No.: 4.88e-14 Length: 68  
Score: 23.00 Matches: 23  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.56% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-966-880a-7 (1-2818) x AAB64608 (1-68)  
  
QY 2054 CTGCACCTCCACCTCCTGGGTTCAAGCATTCTCTGCTCAGCTCCAGCTCCAGTACTCG 1995  
Db 18 LeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrp 37  
  
QY 1994 GATTACAGG 1986  
Db 38 AspTyrArg 40  
  
RESULT 6  
ID ABB11568 standard; peptide: 88 AA.  
XX  
AC ABB11568;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human secreted protein homologue, SEQ ID NO:1938.  
XX  
DE Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; actlyin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumor; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antistimatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytosatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnerary; antitumor.  
XX  
OS Homo sapiens.  
XX  
PN WO200157188-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US03800.  
XX  
PR 03-FEB-2000; 2000US-0496914.  
XX  
PR 27-APR-2000; 2000US-0560875.  
XX

PA (HISE-) HISEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI: 2001-457740/49.  
DR N-PSDB, ABA08812.  
XX  
XX Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
PT e.g. arthritis and cancer -  
XX  
XX Claim 20: Page 213; 1963pp; English.  
XX  
XX Sequences ABB10981-ABB12310 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g. by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention.  
XX  
XX  
SQ Sequence 88 AA:  
  
Alignment Scores:  
Pred. No.: 4.82e-14 Length: 88  
Score: 23.00 Matches: 23  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.56% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-966-880A-7 (1-2818) x ABB11568 (1-88)  
  
QY 2054 CTGCACCTCCACCTCCTGGGTCAAGCATTCCTCGCTCAGCCTCCCAAGTACTGG 1995  
Db 6 LeuGlnProProProProGlyPheLysAlaGpHeSerCysLeuSerLeuProSerSertrp 25  
QY 1994 GATTACAGG 1986  
Db 26 Aspyrarg 28  
  
RESULT 7  
AAU22989

ID AAU22989 standard; Protein; 95 AA.  
XX  
AC AAU22989;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Novel human enzyme polypeptide #75.  
XX  
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
KW nephrotropic; anticoagulant.  
XX  
OS Homo sapiens.  
XX  
PN WO200155301-A2.  
XX  
PD 02-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01239.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225277.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 06-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 23-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249254.  
PR 17-NOV-2000; 2000US-0249255.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PA  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX DR WPI; 2001-465566/50.  
XX DR N-PSDB; AAS40859.  
XX  
PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases  
XX  
XX PS Claim 11; SEQ ID No 985; 1180bp; English.  
XX  
CC The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 95 AA;  
Alignment Scores:  
Pred. No.: 4,8e-14 Length: 95  
Score: 23.00 Matches: 23  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.56% Indels: 0  
DB: 22 Gaps: 0  
US-09-966-880a-7 (1-2818) x AAU22989 (1-95)  
QY 2054 CTGCAACGTCACCTCTGGGTTCAAGGATTCCTCTGCTCAGCCCTCCAGTAGCTG 1995  
Db |||||||  
31 LeuGlnProProProGlyPheIysArgPheSerCysLeuSerLeuProSerSerTrp 50  
QY 1994 GATTACAGG 1986  
Db |||||||  
51 AspTyrArg 53  
RESULT 8  
AAO06720  
ID AAO06720 standard; Protein; 102 AA.  
XX  
AC AAO06720;

```

XX 06-NOV-2001 (first entry)
DE
XX Human polypeptide SEQ ID NO 20612.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukemia;
XX nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX N-PSDB; AA186651.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukemia, inflammation and immune
XX disorders -
XX
XX Claim 20; SEQ ID NO 20612; 1399p + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, hematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukemia, nervous system disorders, arthritis and
XX inflammation.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 102 AA;
XX
XX Alignment Scores:
XX
XX Pred. No.: 4,79e-14 Length: 102
XX
XX Score: 23.00 Matches: 23
XX
XX Percent Similarity: 100.00% Conservative: 0
XX
XX Best Local Similarity: 100.00% Mismatches: 0
XX
XX Query Match: 2,56% Indels: 0
XX
XX DB: 22 Gaps: 0
XX
XX US-09-966-880A-7 (1-2818) x AA006720 (1-102)
XX
XX 2054 CTGCACCTGCACCTCTGGGTTCAAGGATTCCTCTGCCTCAGCCTCCAGTAGCTGG 1995
XX |||||
XX 22 LeuGINProPProPProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrp 41
XX |||||
XX 1994 GATTACAGG 1986
XX |||||
XX 42 AspyrArg 44
XX
XX RESULT 9
XX ID AAG01246
XX ID AAG01246 standard; Protein; 103 AA.
XX

```

AC	AAG01246	
XX	06-OCT-2000	(first entry)
DT		
XX		
DE	Human secreted protein, SEQ ID NO: 5327.	
XX		
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;	
KM	gene therapy; chromosome mapping.	
XX		
OS	Homo sapiens.	
PN	EPI033401.A2.	
XX		
PD	06-SEP-2000.	
XX		
PZ	21-FEB-2000; 2000EP-0200610.	
XX		
PR	26-FEB-1999; 99US-0122487.	
XX		
PA	(GEST ) GENSET.	
XX		
PI	Dumas Milne Edwards J, Duclert A, Giordano J;	
DR	WPI: 2000-500381/45.	
DR	N-PSTDB; AAC01252.	
XX		
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -	
PS	Claim 13; SEQ ID 5327; 71pp + CD-ROM; English.	
XX		
CC	The present sequence is a polypeptide encoded by one of a large number	
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs	
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30	
CC	different tissues. EST sequences usually correspond mainly to the 3'	
CC	untranslated region (UTR) of the mRNA because they are often obtained	
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for	
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in	
CC	those cases where longer cDNA sequences have been obtained, the full 5'	
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'	
CC	ends and can therefore be used to obtain full length cDNAs and genomic	
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and	
CC	chromosome mapping procedures. They are used to obtain upstream	
CC	regulatory sequences and to design expression and secretion vectors.	
XX		
SO	Sequence 103 AA;	
	Alignment Scores:	
	Pred. No.: 4.78e-14 Length: 103	
	Score: 23.00 Matches: 23	
	Percent Similarity: 100.00% Conservative: 0	
	Best Local Similarity: 100.00% Mismatches: 0	
	Query Match: 2.56% Indels: 0	
	DB: 21 Gaps: 0	
	US-09-966-880A-7 (1-2818) x AAG01246 (1-103)	
OY	2054 CTGCACCTCACCACCGGGGTCAAGCATTCCTGCTCCAGCCATCCAAATAGTCG	1995
Db	20 LeuGINPProPProPProGIYPheIysArgPheSerCysLeuSerIeuProSerSerrTP	39
OY	1994 GATTACAGG 1986	
Db	40 AsprYrarg 42	
	RESULT 10	
	AAU31902	
ID	AAU31902 standard; Protein: 107 AA.	
AC	AAU31902;	
XX		
XX		
DT	18-DEC-2001 (first entry)	

XX DE Novel human secreted protein #2393.  
XX  
XX Human; vaccination; gene therapy; nutritional supplement;  
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
XX  
XX Homo sapiens.  
XX PN WO200179449-A2.  
XX PD 25-Oct-2001.  
XX PF 16-Apr-2001; 2001WO-US08656.  
XX PR 18-Apr-2000; 2000US-0552929.  
XX PR 26-Jan-2001; 2001US-0770160.  
XX (HYSE-) HYSEQ INC.  
XX PA Tang YT, Liu C, Drmanac RT;  
XX PI WPI; 2001-611725/70.  
XX DR Nucleic acids encoding a range of human polypeptides, useful in genetic  
XX PT vaccination, testing and therapy -  
XX PS Claim 20; Page 524; 765pp; English.  
XX  
XX The invention relates to novel human secreted polypeptides. The  
XX polypeptides and antibodies to the polypeptides are useful for  
XX determining the presence of or predisposition to a disease associated  
XX with altered levels of polypeptide. The polypeptides are also useful for  
XX identifying agents (agonists and antagonists) that bind to them. Cells  
XX expressing the proteins are useful for identifying a therapeutic agent  
XX for use in treatment of a pathology related to aberrant expression or  
XX physiological interactions of the polypeptide. Vectors comprising  
XX the nucleic acids encoding the polypeptides and cells genetically  
XX engineered to express them are also useful for producing the proteins.  
XX The proteins are useful in genetic vaccination, testing and  
XX therapy, and can be used as nutritional supplements. They may be used to  
XX increase stem cell proliferation; to regulate haematopoiesis; and in  
XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
XX immune suppression and/or stimulation; as anti-inflammatory agents; and  
XX in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid  
XX sequences of novel human secreted proteins of the invention.  
XX  
SQ Sequence 107 AA;  
  
Alignment Scores:  
Pred. No.: 4,78e-14 Length: 107  
Score: 23.00 Matches: 23  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.56% Indels: 0  
DB: Gaps: 0  
US-09-966-880a-7 (1-2818) x AAU31902 (1-107)  
  
QY 2054 CTGCACCTCCACCTCTCTGGGTTCAAGCATTTCTCTGCTCAGCCTCCCAAGTACG 1995  
Db 29 LeuGlnProProProGlyPheIysArgPheSerCysIeuSerIeuProSerSerTrp 48  
QY 1994 GATTACAGG 1986  
Db 49 AspyrArg 51  
  
RESULT 11  
AA002085  
ID AA002085 standard; Protein: 117 AA.  
XX AA002085;  
XX

DT 06-NOV-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 15977.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX OS  
XX PN WO200164835-A2.  
XX PD 07-SEP-2001.  
XX PF 26-FEB-2001; 2001WO-US04927.  
XX PR 28-FEB-2000; 2000US-0515126.  
XX PR 18-May-2000; 2000US-0577409.  
XX (HYSE-) HYSEQ INC.  
XX PA Tang YT, Liu C, Drmanac RT;  
XX PI WPI; 2001-514838/56.  
XX DR N-PSDB; AA182016.  
XX DR N-PSDB; AA182016.  
XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX  
XX Claim 20; SEQ ID NO 15977; 1399p + Sequence listing; English.  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity inducing to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activity/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_Pct\_sequences.  
XX  
SQ Sequence 117 AA;  
  
Alignment Scores:  
Pred. No.: 4,75e-14 Length: 117  
Score: 23.00 Matches: 23  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.56% Indels: 0  
DB: Gaps: 0  
US-09-966-880a-7 (1-2818) x AA002085 (1-117)  
  
QY 2054 CTGCACCTCCACCTCTCTGGGTTCAAGCATTTCTCTGCTCAGCCTCCCAAGTACG 1995  
Db 80 LeuGlnProProProGlyPheIysArgPheSerCysIeuSerIeuProSerSerTrp 99  
QY 1994 GATTACAGG 1986  
Db 100 AspyrArg 102  
  
RESULT 12  
ABG12796  
ID ABG12796 standard; Protein: 211 AA.  
XX ABG12796;  
XX

XX 18-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #12787.  
 DE Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HSE-) HXSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB: AAS76983.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 43155; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 211 AA:  
 Alignment Scores:  
 Pred. No.: 4,62e-14 Length: 211  
 Score: 23.00 Matches: 23  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.56% Indels: 0  
 DB: Gaps: 0  
 US-09-966-880A-7 (1-2818) x ABG12796 (1-211)  
 QY 2054 CTGCAACCTCCACCTCTCTGGGTTCAAGCATTTCTCTGCTCCAGCTCCCAAGTAGCTGG 1995  
 Db 159 TleuglProbrProbrProbrGlyPheLysArgPheSerCysLeuSerLeuProserSertrp 178  
 QY 1994 GATTACAGG 1986  
 |||||||||

DB 179 AspyrArg 181  
 RESULT 13  
 ABG22078  
 ID ABG22078 standard; Protein: 264 AA.  
 XX  
 AC ABG22078;  
 XX  
 DI 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #22069.  
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HSE-) HXSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB: AAS86265.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 52437; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 264 AA:  
 Alignment Scores:  
 Pred. No.: 4,81e-13 Length: 264  
 Score: 22.00 Matches: 22  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.45% Indels: 0  
 DB: Gaps: 0  
 US-09-966-880A-7 (1-2818) x ABG22078 (1-264)



US-09-966-880A-7 (1-2818) x AAU31820 (1-264)

OY 2054 CTGCAACCTCCACCTCCTGGGCTTCAAGCGAATTCCTCCCTCAGCCTCCAAAGTAGCTGG 1999S  
Db 139 LeuGlnProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrp 158  
OY 1994 GATTAC 1989  
Db 159 AspIyr 160

RESULT 15  
AA009322  
ID AA009322 standard; Protein; 32 AA.  
XX  
AC AA009322;  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 23214.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
PN WO200164835-A2.  
PD 07-SEP-2001.  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
PA (HYSEQ-) HYSEQ INC.  
PI Tang YT, Liu C, Drmanac RT;  
PI  
DR WPI: 2001-514838/56.  
DR N-PSDB; AAI189253.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 23214; 1399bp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibitin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/Published\_pct\_sequences.  
XX

SQ	sequence	32	AA:	
Alignment Scores:				
Pred. No.:	5.6e-12	Length:	32	
Score:	21.00	Matches:	21	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	2.34%	Indels:	0	

US-09-966-880A-7 (1-2818) x AA009322 (1-32)

QY 2054 CTGCAACCTCCACACCTCTGCGTTCAAGCGATTCTCTGCGCTCAGCCTCCCAAGTAGCTGG 1995

Db 10 LeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuProSerTrp 29

QY 1994 GAT 1992

Db 30 Asp 30

Search completed: July 7, 2003, 23:30:46  
 Job time : 78.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: July 7, 2003, 23:28:07 ; Search time 27 Seconds  
(without alignments)  
6141.763 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 897  
Sequence: 1 agaggaaccatcattatgta.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table:

	OLIGO	Xgapop	Xgapext
	60.0	60.0	60.0
	Ygapop	60.0	60.0
	Fgapop	6.0	7.0
	Delop	6.0	7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+.n2p.model -DEV=xip  
-Q=/cgn2\_1/USPTO.spool/US09966880/runat.07072003.142234.22114/app.query.fasta\_1.3015  
-DB=Issued.Patents.AA -OPMT=fastan -SUFFIX=olig.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR.SCORE=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFM=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09966880 -CGN\_1.1\_22\_@runat.07072003.142234.22114 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

Issued.Patents.AA.\*  
1: /cgn2\_6/ptodata/1/aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/aa/backfilest1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 2	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 3	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 4	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 5	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 6	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 7	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 8	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 9	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 10	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 11	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App
C 12	16	1.8	375	2	US-08-454-557C-121 Sequence 121, App

13	10	1.1	49	4	US-09-288-143-93	Sequence 93, Appl
14	10	1.1	65	4	US-09-227-357-538	Sequence 538, App
15	10	1.1	75	2	US-08-454-557C-48	Sequence 48, Appl
16	10	1.1	75	2	US-08-340-426D-48	Sequence 48, Appl
17	10	1.1	75	2	US-08-450-673C-48	Sequence 48, Appl
18	10	1.1	75	5	PCT-US95-17111A-48	Sequence 48, Appl
19	10	1.1	312	4	US-09-254-465A-9	Sequence 9, Appl
20	10	1.1	388	4	US-09-265-630-11	Sequence 11, Appl
21	10	1.1	416	2	US-08-694-915-4	Sequence 13, Appl
22	10	1.1	500	4	US-09-265-630-13	Sequence 13, Appl
23	10	1.1	510	4	US-09-211-417-1	Sequence 14, App
24	9	1.0	18	3	US-08-851-843A-214	Sequence 214, App
25	9	1.0	18	4	US-08-974-549A-333	Sequence 333, App
26	9	1.0	18	4	US-08-854-050-214	Sequence 214, App
27	9	1.0	18	4	US-09-430-332-214	Sequence 214, App
28	9	1.0	56	4	US-09-227-357-537	Sequence 377, App
29	9	1.0	57	4	US-09-082-358B-43	Sequence 43, Appl
30	9	1.0	96	4	US-09-605-785-827	Sequence 827, App
31	9	1.0	128	4	US-09-342-847A-4	Sequence 4, Appl
32	9	1.0	132	1	US-08-534-975-4	Sequence 4, Appl
33	9	1.0	132	2	US-08-954-470-4	Sequence 4, Appl
34	9	1.0	132	3	US-09-129-855A-4	Sequence 4, Appl
35	9	1.0	132	4	US-09-247-154-4	Sequence 4, Appl
36	9	1.0	132	4	US-09-480-718-4	Sequence 4, Appl
37	9	1.0	224	4	US-09-605-785-825	Sequence 825, App
38	9	1.0	255	4	US-09-227-357-488	Sequence 488, App
39	9	1.0	1261	4	US-09-605-785-538	Sequence 538, App
40	9	1.0	1261	4	US-09-439-313-538	Sequence 7, Appl
41	9	1.0	1497	3	US-08-623-679-7	Sequence 7, Appl
42	9	1.0	1497	3	US-08-933-774-7	Sequence 7, Appl
43	9	1.0	1497	4	US-09-181-030-7	Sequence 7, Appl
44	9	1.0	1497	4	US-09-534-242-7	Sequence 7, Appl
45	9	1.0	1497	4	US-09-454-854-7	Sequence 7, Appl

#### ALIGNMENTS

RESULT 1  
US-08-454-557C-121  
; Sequence 121, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; TITLE OF INVENTION: Neutral Thread Protein Gene Expression and Detection  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454,557C  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609,3840003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 121:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 375 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-454-557C-121

## Alignment Scores:

Pred. No.:	1.12e-06	Length:	375
Score:	16.00	Matches:	16
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.78%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-7 (1-2818) x US-08-454-557C-121 (1-375)

OY 1879 TCGGCTCCCAAGTGTGGATTACAGCGGTGAGCCACGACCCCG 1832

DB 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaArg 294

## RESULT 2

US-08-340-426D-121

Sequence 121, Application US/08340426D

Patent No. 5948634

GENERAL INFORMATION:

APPLICANT: de la Monte, Suzanne

TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

NUMBER OF SEQUENCES: 121

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/340,426D

FILING DATE: 14-NOV-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, Steven R.

REGISTRATION NUMBER: 36,203

REFERENCE/DOCKET NUMBER: 0609.3840002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 121:

SEQUENCE CHARACTERISTICS:

LENGTH: 375 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-340-426D-121

Alignment Scores:

Pred. No.:	1.12e-06	Length:	375
Score:	16.00	Matches:	16
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.78%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-7 (1-2818) x US-08-340-426D-121 (1-375)

OY 1879 TCGGCTCCCAAGTGTGGATTACAGCGGTGAGCCACGACCCCG 1832

DB 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaArg 294

## RESULT 3

US-08-450-673C-121

Sequence 121, Application US/08450673C

Patent No. 5948888

GENERAL INFORMATION:

APPLICANT: de la Monte, Suzanne

TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

NUMBER OF SEQUENCES: 121

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450,673C

FILING DATE: 30-MAY-1995

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, Steven R.

REGISTRATION NUMBER: 36,203

REFERENCE/DOCKET NUMBER: 0609.3840004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 121:

SEQUENCE CHARACTERISTICS:

LENGTH: 375 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-450-673C-121

Alignment Scores:

Pred. No.:	1.12e-06	Length:	375
Score:	16.00	Matches:	16
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.78%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-7 (1-2818) x US-08-450-673C-121 (1-375)

OY 1879 TCGGCTCCCAAGTGTGGATTACAGCGGTGAGCCACGACCCCG 1832

DB 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaArg 294

## RESULT 4

US-09-058-489-22

Sequence 22, Application US/09058489

Patent No. 6103886

GENERAL INFORMATION:

APPLICANT: Whitehead Institute for Biomedical Research

APPLICANT: Lahn, Bruce

TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

FILE REFERENCE: WH:97-088A

CURRENT APPLICATION NUMBER: US/09/058,489

EARLIER FILING DATE: 1998-04-10

EARLIER APPLICATION NUMBER: 60/041,877

NUMBER OF SEQ ID NOS: 91

SOFTWARE: FastSeq for Windows Version 3.0

```
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:
Pred. No.:          9,96e-07      Length:      1079
Score:             16.00          Matches:      16
Percent Similarity: 100.00%       Conservative: 0
Best Local Similarity: 100.00%    Mismatches:  0
Query Match:       1.78%          Indels:      0
DB:                3              Gaps:        0

US-09-966-880a-7 (1-2818) x US-09-058-489-22 (1-1079)

QY      1879 TCGGCTCCCAAGTGTGGATTACAGCGGTGAGCCACGCGCCG 1832
        |||||||
Db       1064 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaArg 1079

RESULT 5
US-09-227-357-171
; Sequence 171, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13664
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953

; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (90)
; OTHER INFORMATION: Xaa equals stop translation
US-09-227-357-171

Alignment Scores:
Pred. No.:          1.15e-05      Length:      90
Score:             15.00          Matches:      15
Percent Similarity: 100.00%       Conservative:  0
Best Local Similarity: 100.00%    Mismatches:  0
Query Match:       1.67%          Indels:      0
DB:                4              Gaps:        0

US-09-966-880a-7 (1-2818) x US-09-227-357-171 (1-90)

QY      1879 TCGGCTCCCAAGTGTGGATTACAGCGGTGAGCCACGCGCC 1835
        |||||||
Db       36  SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAla 50

RESULT 6
PCT-US95-17111A-121
; Sequence 121, Application PC/TUS951711A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neutral Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
```

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/340,426  
FILING DATE: 14-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0609.3840002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-17111A-121

Alignment Scores:  
Pred. No.: 8.62e-05 Length: 397  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.56% Indels: 0  
Gaps: 5

US-09-966-880A-7 (1-2818) x PCT-US95-17111A-121 (1-397)

QY 2030 AACGATTCCTGCTGCTCAGCTCCAGTAGCTGGATTAC 1989  
|||||  
Db 331 LysArgPheSerCysLeuSerLeuProSerSerTTPaPtyr 344

RESULT 7  
US-08-704-711A-1  
Sequence 1, Application US/08704711A  
Patent No. 6114159  
GENERAL INFORMATION:  
APPLICANT: WILL, Horst  
APPLICANT: HINZMANN, Bernd  
TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX  
METALLOPROTEASES, THEIR PRODUCTION AND USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,711A  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/DE95/00357  
FILING DATE: 17-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 4438838.1  
FILING DATE: 21-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 4409663.1  
FILING DATE: 17-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GRANADOS, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 26083/124  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 579 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-704-711A-1

Alignment Scores:  
Pred. No.: 0.00073 Length: 579  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.45% Indels: 0  
Gaps: 3

US-09-966-880A-7 (1-2818) x US-08-704-711A-1 (1-579)

QY 2041 CTCCTGGATTCAACGATTCCTGCTGCTCAGCTCCCAA 2003  
|||||  
Db 60 LeuEnglySerSerAspSerProAlaSerAlaSerGln 72

RESULT 8  
US-09-521-220-1  
Sequence 1, Application US/09521220  
Patent No. 639348  
GENERAL INFORMATION:  
APPLICANT: WILL, Horst  
APPLICANT: HINZMANN, Bernd  
TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX  
METALLOPROTEASES, THEIR PRODUCTION AND USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/521,220  
FILING DATE: 08-MAR-2000  
CLASSIFICATION: <Unknown>  
21-OCT-1994  
17-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/704,711  
FILING DATE: <Unknown>  
APPLICATION NUMBER: DE 4438838.1  
FILING DATE: 21-OCT-1994  
APPLICATION NUMBER: DE 4409663.1  
FILING DATE: 17-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GRANADOS, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 26083/124  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 579 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

```

;
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-521-220-1

Alignment Scores:
Pred. No.: 0.00073 Length: 579
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.45% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-521-220-1 (1-579)

QY 2041 CTCCTGGTTCAGCAGGCTTCTCTGCTCAGCTCCCA 2003
DB 60 LeuLenglyserSersPserProAlaSerAlaSerGln 72

RESULT 9
US-09-288-143-170
; Sequence 170, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; EARLIER FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-288-143-170

Alignment Scores:
Pred. No.: 0.0766 Length: 35
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.22% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-288-143-170 (1-35)

QY 1933 TGGCAGTGTGGCAGGCTGCTCAACTCTCG 1901
DB 24 CysH1sValaIglyInAlaGlyLeuLysLeuLeu 34

RESULT 10
US-09-288-143-168
; Sequence 168, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; EARLIER FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-288-143-168

Alignment Scores:
Pred. No.: 0.0685 Length: 99
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.22% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-288-143-168 (1-99)

QY 1933 TSCCATGTTGGCAGCAGGCTGCTCAACTCTCG 1901
DB 56 CysH1sValaIglyInAlaGlyLeuLysLeuLeu 66

RESULT 11
US-09-227-357-285
; Sequence 285, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; EARLIER FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
```

```

; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 285
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-285

Alignment Scores:
Pred. No.: 0.067      Length: 122
Score: 11.00         Matches: 11
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.22%      Indels: 0
Gaps: 0
DB: 4

US-09-966-880a-7 (1-2818) x US-09-227-357-285 (1-122)
QY 1930 CAGTTGGCCAGGCTGATCTCAAACTCTGACC 1898
Db 42 HisValGlyInAlaGlyLeuLysLeuLeuThr 52

RESULT 12
US-09-288-143-171
; Sequence 171, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
```

```

; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 171
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-288-143-171

Alignment Scores:
Pred. No.: 0.682      Length: 32
Score: 10.00         Matches: 10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.11%      Indels: 0
Gaps: 0
DB: 4

US-09-966-880a-7 (1-2818) x US-09-288-143-171 (1-32)
QY 1879 TCGGCTCCCAAGTCGGATACAGCC 1850
Db 7 SerAlaSerGlnSerAlaGlyLeuThrCly 16

RESULT 13
US-09-288-143-93
; Sequence 93, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 93
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-288-143-93

Alignment Scores:
Pred. No.: 0.651      Length: 49
Score: 10.00         Matches: 10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.11%      Indels: 0
Gaps: 0
```



DB: 4 Gaps: 0  
US-09-966-880a-7 (1-2818) x US-09-288-143-93 (1-49)  
QY 1842 TGCGTACGCCCTGATCCAGCACTTGG 1871  
Db 16 TpleuthrProvalleProAlaLeutrip 25  
RESULT 14  
US-09-227-357-538  
; Sequence 538, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,916  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,930  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,918  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,920  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,733  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,795  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,919  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,928  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/055,722  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,723  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,948  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,949  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,953  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,950  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,947  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,964  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/056,360  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,684  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,984

; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,954  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/058,785  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,664  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,660  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,661  
; EARLIER FILING DATE: 1997-09-12  
; NUMBER OF SEQ ID NOS: 672  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 538  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-227-357-538  
Alignment Scores:  
Pred. No.: 0.632 Length: 65  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.11% Indels: 0  
Gaps: 0  
DB: 4 Gaps: 0  
US-09-966-880a-7 (1-2818) x US-09-227-357-538 (1-65)  
QY 1851 CCTGATCCGACACTTGGAGCCGAG 1880  
Db 43 ProvalleProAlaLeutripAlaIn 52  
RESULT 15  
US-08-454-557C-48  
; Sequence 48, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454,557C  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609,3840003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 75 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-454-557C-48

Alignment Scores:  
 Pred. No.: 0.623 Length: 75  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.11% Indels: 0  
 DB: 2 Gaps: 0

US-09-966-880a-7 (1-2818) x US-08-454-557C-48 (1-75)

QY 1880 CTGGCCCTCCAAAGTGCTGGGATTACAGG 1851

Db 4 LeuGlyLeuProLysCysTrpAspTyrArg 13

Search completed: July 7, 2003, 23:40:00  
 Job time : 35 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - protein search, using frame\_plus\_m2p model

Run on: July 7, 2003, 23:30:52 ; Search time 69 Seconds  
(without alignments)  
9392.755 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 897  
Sequence: 1 agagaaccatcattatga.....aaaaaaaaaaaaaaaaa 2818

Scoring table:  
OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 440863 seqs, 114932915 residues

Word size: 1

Total number of hits satisfying chosen parameters: 826674

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-O=/cgn2\_1/USPRO.spool/US09966880/runat\_07072003\_142234\_22138/app.Query.fasta\_1.3015  
-DB=Published\_Applications\_MA-QFMT-tastan-SUFFIX=olig.rapb-KINMATCH=0.1  
-LOOPEXT=0-LOOPEXT=0-UNITS-bits-START=1-END=1-MATRIX=oligo  
-TRANS-human40.cdi-LIST=45-DOCALIGN=200-THR\_SCORE=quality-THR\_MIN=1  
-ALIGN=15-MODE=LOCAL-OUTFMT=ptc-NORM-ext-HEAPSIZE=500-MINLEN=0  
-MAXLEN=2000000000-USER=US09966880@cgn.1.1.86@runat.07072003\_142234\_22138  
-NCPU=6-ICPU=3-NO\_MAP-LARGEQUERY-NEG\_SCORES=0-NAIT-DSBIOCK=100  
-LONLOG-DEV\_TIMEOUT=120-WARN\_TIMEOUT=30-THREADS=1-XGAPOP=60-XGAPEXT=60  
-FGAPOP=6-FGAPEXT=7-YGAPOP=60-YGAPEXT=60-DELOP=6-DELEXT=7

Database: Published\_Applications\_MA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	198	22.1	198	9	US-09-966-880A-8
2	42	4.7	198	9	US-09-966-880A-2
3	19	2.1	59	9	US-10-106-698-5145
4	18	2.0	44	10	US-09-925-302-803

5	18	2.0	86	9	US-10-011-585A-154	Sequence 154, App
6	18	2.0	147	9	US-09-764-891-3556	Sequence 3556, App
7	18	2.0	368	9	US-10-198-070-79	Sequence 79, Appl
8	17	1.9	51	9	US-09-764-891-4746	Sequence 4746, Ap
9	17	1.9	90	9	US-10-001-873-34	Sequence 34, Appl
10	17	1.9	133	9	US-10-198-070-45	Sequence 45, Appl
11	16	1.8	25	9	US-09-925-299-1013	Sequence 1013, Ap
12	16	1.8	25	10	US-09-925-299-1013	Sequence 1013, Ap
13	16	1.8	85	10	US-09-925-297-829	Sequence 829, App
14	16	1.8	91	9	US-10-023-282-781	Sequence 781, App
15	16	1.8	155	9	US-10-198-070-24	Sequence 24, Appl
16	16	1.8	181	9	US-10-198-070-114	Sequence 114, App
17	16	1.8	232	9	US-10-198-070-34	Sequence 34, Appl
18	16	1.8	375	9	US-10-146-130-2	Sequence 2, Appl1
19	16	1.8	375	9	US-09-964-667-2	Sequence 2, Appl1
20	16	1.8	375	9	US-09-872-968-2	Sequence 2, Appl1
21	16	1.8	375	9	US-10-092-934-10	Sequence 10, Appl
22	16	1.8	375	9	US-10-153-334-1	Sequence 1, Appl1
23	16	1.8	375	9	US-10-198-069-1	Sequence 1, Appl1
24	16	1.8	375	9	US-10-157-031-299	Sequence 299, App
25	16	1.8	375	9	US-10-198-070-1	Sequence 1, Appl1
26	16	1.8	375	9	US-09-964-666-2	Sequence 2, Appl1
27	16	1.8	375	10	US-09-964-412-2	Sequence 2, Appl1
28	15	1.7	63	9	US-09-764-891-3498	Sequence 3498, App
29	15	1.7	70	9	US-09-764-872-288	Sequence 288, App
30	15	1.7	71	9	US-09-925-299-1026	Sequence 1026, Ap
31	15	1.7	71	10	US-09-925-299-1026	Sequence 1026, Ap
32	15	1.7	75	9	US-10-091-504-1151	Sequence 1151, Ap
33	15	1.7	75	10	US-09-764-869-1151	Sequence 1151, Ap
34	15	1.7	90	9	US-09-983-802-171	Sequence 171, App
35	15	1.7	181	9	US-10-198-070-55	Sequence 55, Appl
36	15	1.6	19	9	US-10-153-334-49	Sequence 49, Appl
37	14	1.6	46	9	US-10-198-069-12	Sequence 12, Appl
38	14	1.6	46	9	US-10-153-334-48	Sequence 48, Appl
39	14	1.6	58	9	US-09-764-891-4323	Sequence 4323, Ap
40	14	1.6	62	9	US-10-106-698-4816	Sequence 4816, App
41	14	1.6	76	9	US-10-198-070-107	Sequence 107, App
42	14	1.6	94	9	US-10-001-833-135	Sequence 135, App
43	14	1.6	100	9	US-09-989-920-264	Sequence 264, App
44	14	1.6	115	10	US-09-995-494-79	Sequence 79, Appl
45	14	1.6	118	9	US-10-198-070-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1  
US-09-966-880A-8  
Sequence 8, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Muramatsu Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 198  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-966-880A-8  
Alignment Scores:



```

CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3556
LENGTH: 147
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (134)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (138)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (140)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3556

Alignment Scores:
Pred. No.:      1.16e-06      Length:      147
Score:          18.00        Matches:     18
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    2.00%       Indels:      0
DB:             9           Gaps:       0

US-09-966-880A-7 (1-2818) x US-09-764-891-3556 (1-147)
QY      2042 CCTCGTGGATTCAAGGATTCCTGCTCAGCCCTCCCAAGTAGCTGGATTAC 1989
Db      96 ProProgl1phelysarphesercyscysleuPanserleuProserSerrTPAspyr 113
|||||
RESULT 7
US-10-198-070-79
Sequence 79, Application US/10198070
Publication No.: US20030109437A1
GENERAL INFORMATION:
APPLICANT: AVERBACK, PAUL
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
FILE REFERENCE: 59003.000008
CURRENT APPLICATION NUMBER: US/10/198,070
CURRENT FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/306,161
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/306,150
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/331,477
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 79
LENGTH: 368
TYPE: PRT
ORGANISM: Homo sapiens
US-10-198-070-79

Alignment Scores:
Pred. No.:      9.71e-07      Length:      368
Score:          18.00        Matches:     18
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    2.00%       Indels:      0
DB:             9           Gaps:       0

US-09-966-880A-7 (1-2818) x US-10-198-070-79 (1-368)
QY      2039 CCGGTTCAGCGATTCTCTGCTCAGCCCTCCCAAGTAGCTGGATTACAGG 1986

```

Db 25 ProGlyPheLysArgPheSerCysLeuSerLeuProSerSertPaspTyrArg 42

## RESULT 8

US-09-764-891-4746  
; Sequence 4746, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4746  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (4)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-4746

## Alignment Scores:

Pred. No.: 1 03e-05 Length: 51  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.90% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-764-891-4746 (1-51)

QY 1993 TCCGACGACTGTTGGAGCTGAGCGAGAGATCGCTTGACCCAGAGGT 2043

Db 17 SerGlnLeuLengLylArgLeuArgGlnGlnAsnArgLeuAsnProGlyGly 33

## RESULT 9

US-10-001-873-34  
; Sequence 34, Application US/10001873  
; Patent No. US20020160388A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Turner, Leah  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot  
; FILE REFERENCE: DEX-0275  
; CURRENT APPLICATION NUMBER: US/10/001,873  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,055  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: 60/252,496  
; PRIOR FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34  
; LENGTH: 90  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-001-873-34

## Alignment Scores:

Pred. No.: 9 22e-06 Length: 90  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.89% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-10-001-873-34 (1-90)

QY 2032 TCACGCAATCTCTGCTGCTAGCCTCCAGTACTGCTGGATTACAGTGCC 1982

Db 28 SerSerAspSerProAlaSerAlaSerGlnValAlaGlyIleThrGlyAla 44

## RESULT 10

US-10-198-070-45  
; Sequence 45, Application US/10198070  
; Publication No. US20030109437A1  
; GENERAL INFORMATION:  
; APPLICANT: AVERBACK, PAUL  
; APPLICANT: GEMMELL, JACK  
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 59003,000008  
; CURRENT APPLICATION NUMBER: US/10/198,070  
; CURRENT FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 60/306,161  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/306,150  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/331,477  
; PRIOR FILING DATE: 2001-11-16  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 45  
; LENGTH: 133  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-198-070-45

## Alignment Scores:

Pred. No.: 8 53e-06 Length: 133  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.89% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-10-198-070-45 (1-133)

QY 2041 CTCCTGGCTTACCGATTCCTGCTGAGCTCCCACTAGCTGGAGT 1991

Db 92 LeuLengLysSerSerProAlaSerAlaSerGlnValAlaGlyIle 108

## RESULT 11

US-09-925-299-1013  
; Sequence 1013, Application US/09925299  
; Publication No. US20030040617A3  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1013  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-299-1013

## Alignment Scores:

Pred. No.: 8 53e-05 Length: 25  
Score: 16.00 Matches: 16

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.78% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-925-299-1013 (1-25)

OY 1832 CCGGGCGTGGTGCACGCGCTGTATCCAGCACTTTGGAGGCCGA 1879

DB 5 ProglyvalvalaAlaHisAlaCysAsnProSerThrlleuclgylarg 20

## RESULT 12

US-09-925-299-1013  
; Sequence 1013, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 1013  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-299-1013

## Alignment Scores:

Pred. No.: 8,53e-05 Length: 25  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.78% Indels: 0  
DB: 10 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-925-299-1013 (1-25)

OY 1832 CCGGGCGTGGTGCACGCGCTGTATCCAGCACTTTGGAGGCCGA 1879

DB 5 ProglyvalvalaAlaHisAlaCysAsnProSerThrlleuclgylarg 20

## RESULT 13

US-09-925-297-829  
; Sequence 829, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 829  
; LENGTH: 85  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-297-829

## Alignment Scores:

Pred. No.: 6,7e-05 Length: 85  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.78% Indels: 0  
DB: 10 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-925-297-829 (1-85)

OY 1832 CCGGGCGTGGTGCACGCGCTGTATCCAGCACTTTGGAGGCCGA 1879

DB 53 ProglyvalvalaAlaHisAlaCysAsnProSerThrlleuclgylarg 68

## RESULT 14

US-10-023-282-781  
; Sequence 781, Application US/10023282  
; Publication No. US20030092893A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P1  
; CURRENT APPLICATION NUMBER: US/10/023,282  
; CURRENT FILING DATE: 2001-12-20  
; EARLIER APPLICATION NUMBER: 09/205,258  
; EARLIER FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,915  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875

```

FILE REFERENCE: 59003 000008
CURRENT APPLICATION NUMBER: US/10/198, 070
CURRENT FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/306,161
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/306,150
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/331,477
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 155
TYPE: PRT
ORGANISM: Homo sapiens
US-10-198-070-24

Alignment Scores:
Pred. No.: 5,95e-05      Length: 155
Score: 16.00           Matches: 16
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.78%              Indels: 0
DB: 9                      Gaps: 0

US-09-966-880A-7 (1-2818) x US-10-198-070-24 (1-155)

QY      2032 TCAGGCGATTCCTCGCTCAGCCGCCCAAGTAGCTGGGATTCACAGT 19855
|||||
Db      67  serSerAspSerProAlaSerAlaSerGlnValAlaGlyIleIleThrGly 82

Search completed: July 7, 2003, 23:42:27
Job time : 72 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 7, 2003, 23:19:47; Search time 55.5 Seconds  
(without alignments)  
9762.407 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 897  
Sequence: 1 agaggaaccctcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table:  
OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09966880/rnat.07072003.142233.22095/app.query.fasta.1.3015  
-DB=PIR.73 -OPMT=fastan -SUFFIX=olig.rpr -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.coi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pio  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US09966880.qcgn.1.1.72.qunat.07072003.142233.22095 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

PIR.73:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1.7	627	4	A40201	artifact-warning s
C 2	1.6	613	4	C40201	artifact-warning s
C 3	1.4	296	2	S01441	class II histocomp
C 4	1.3	597	4	E40201	artifact-warning s
C 5	1.3	613	4	C40201	artifact-warning s
C 6	1.2	673	4	F40201	artifact-warning s
C 7	1.3	841	1	I78885	serine/threonine-s
C 8	1.2	47	2	I79424	MHC HLA-DR-beta-1
C 9	1.2	114	2	JC5238	galactosylceramide
C 10	1.1	301	4	B40201	artifact-warning s
C 11	1.2	391	2	T17239	hypothetical prote
C 12	1.1	597	4	E40201	artifact-warning s
C 13	1.2	627	4	A40201	artifact-warning s
C 14	1.2	673	4	F40201	artifact-warning s

C 15	1.1	1.2	773	2	T00502	probable receptor-
C 16	1.0	1.1	200	2	A6010	X-linked retinopat
C 17	1.0	1.1	252	2	PC4259	ferritin associate
C 18	1.0	1.1	832	2	T49494	condensin complex
C 19	1.0	1.1	2010	2	B71616	phosphatase (acid
C 20	1.0	1.1	2708	2	T09079	probable chloroqui
C 21	1.0	1.1	2819	2	T09080	probable chloroqui
C 22	1.0	1.0	115	2	S69849	hypothetical prote
C 23	1.0	1.0	132	2	T19004	hypothetical prote
C 24	1.0	1.0	166	2	T18513	cyclin-dependent k
C 25	1.0	1.0	167	2	S38112	hypothetical prote
C 26	1.0	1.0	215	2	T52523	hypothetical prote
C 27	1.0	1.0	244	2	T62655	nucleoporin p62 ho
C 28	1.0	1.0	285	2	S29368	hypothetical prote
C 29	1.0	1.0	286	2	A63482	translational initia
C 30	1.0	1.0	328	2	E88930	probable transcrip
C 31	1.0	1.0	333	2	T52355	protein R1G11.11
C 32	1.0	1.0	401	2	T02958	hypothetical prote
C 33	1.0	1.0	432	2	T02958	ribulose-bisphosph
C 34	1.0	1.0	579	4	D40201	MAR16 homolog PFB0
C 35	1.0	1.0	580	2	T18439	artifact-warning s
C 36	1.0	1.0	600	2	T18446	hypothetical prote
C 37	1.0	1.0	635	2	F71621	hypothetical prote
C 38	1.0	1.0	671	2	A40692	hypothetical prote
C 39	1.0	1.0	1263	2	T13805	signal recognition
C 40	1.0	1.0	1791	2	T24089	spalt-related prot
C 41	1.0	1.0	1888	2	T39009	hypothetical prote
C 42	1.0	1.0	2573	2	D71614	hypothetical prote
C 43	1.0	1.0	4550	2	T18440	hypothetical prote
C 44	0.9	0.9	39	2	I54374	gene NF2 protein -
C 45	0.9	0.9	41	2	T07274	hypothetical prote

#### ALIGNMENTS

RESULT 1  
A40201  
artifact-warning sequence (translated ALU class A) - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:Accession: A40201  
R:Claverie, J.M.  
Personal communication, 1992  
A:Reference number: A40201  
A:Accession: A40201  
A:Molecule type: DNA  
A:Residues: 1-627 <CLA>  
R:Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti  
A:Reference number: A40200; MUID:92241891; PMID:1572661  
A:Contents: annotation  
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o  
in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:  
Pred. No.: 3.02e-06  
Score: 15.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 1.67%  
DB: 4  
Gaps: 0  
US-09-966-880A-7 (1-2818) x A40201 (1-627)  
OY 1889 TCCGCCGAGCTCGCCCTCCCAAGTGTGGATTACAGCGMGAG 1845  
|||||  
Db 504 Serralaargleugllyleuprolscysctrrpasprrytrargatgglu 518  
RESULT 2  
C40201

artifact-warning sequence (translated ALU class C) - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:Accession: C40201  
R:Claverie, J.M.  
Personal communication, 1992  
A:Reference number: A40201  
A:Accession: C40201  
A:Molecule type: DNA  
A:Residues: 1-613 <CLAU>  
R:Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential  
A:Reference number: A40200; MUID:92241891; PMID:1572661  
A:Contents: annotation  
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of  
in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:  
Pred. No.: 3.3e-05 Length: 613  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.56% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x C40201 (1-613)

QY 1898 GGTCAGAGCTTGAGACGAGCTGGCCAGCATGGCAAAACC 1939  
Db 23 GlyGInGluPhelGluThrSerLeuAlaAsnMetAlaLysPro 36

RESULT 3  
S01441  
Class II histocompatibility antigen HLA-DR beta-1 chain precursor (clone L51.1.1) - huma  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 21-Jan-2000  
C:Accession: S01441  
R:Catlins, J.S.; Dahl, C.A.; Curijsinger, J.M.; Bach, F.H.  
Nucleic Acids Res. 16, 9353, 1988  
A:Title: Identification of a novel DR beta cDNA clone.  
A:Reference number: S01441; MUID:89016649; PMID:3174462  
A:Accession: S01441  
A:Molecule type: mRNA  
A:Residues: 1-296 <CA1>  
A:Cross-references: EMBL:X12544; NID:932210; PID:CAA31061.1; PID:932211  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1988  
A:Note: only part of the sequence is given  
A:Note: the authors translated the codon TTT for residue 266 as Gly  
C:Superfamily: class II histocompatibility antigen; immunoglobulin homology  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-296/Product: class II histocompatibility antigen, HLA-DR beta-1 chain #status predi  
F:139-204/Domain: immunoglobulin homology <IMM>

Alignment Scores:  
Pred. No.: 0.000386 Length: 296  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.45% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-7 (1-2818) x S01441 (1-296)

QY 1842 TGGCTCAGCCCTGTATCCAGACATTTGGAGGCCGAG 1880  
Db 280 TrpLeuThrProValIleProAlaLeuTrpGlnAlaGln 292

RESULT 4  
E40201  
artifact-warning sequence (translated ALU class E) - human

C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:Accession: E40201  
R:Claverie, J.M.  
Personal communication, 1992  
A:Reference number: A40201  
A:Accession: E40201  
A:Molecule type: DNA  
A:Residues: 1-597 <CLAU>  
R:Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti  
A:Reference number: A40200; MUID:92241891; PMID:1572661  
A:Contents: annotation  
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o  
in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:  
Pred. No.: 0.00394 Length: 597  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.34% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x E40201 (1-597)

QY 2021 TCTGCTCAGCCCTGCCAGAGCTGGATTACAG 1986  
Db 534 SerCysLeuSerLeuProSerSerTrpAspTyrArg 545

RESULT 5  
C40201  
artifact-warning sequence (translated ALU class C) - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:Accession: C40201  
R:Claverie, J.M.  
Personal communication, 1992  
A:Reference number: A40201  
A:Accession: C40201  
A:Molecule type: DNA  
A:Residues: 1-613 <CLAU>  
R:Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti  
A:Reference number: A40200; MUID:92241891; PMID:1572661  
A:Contents: annotation  
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o  
in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:  
Pred. No.: 0.00393 Length: 613  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.34% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x C40201 (1-613)

QY 1938 GCTTTGCCATGTTGGCCAGGCTGCTCAAACTCC 1903  
Db 579 GlyPheAlaMetLeuAlaArgLeuValSerAsnSer 590

RESULT 6  
F40201  
artifact-warning sequence (translated ALU class F) - human  
C:Species: Homo sapiens (man)

C:/Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:/Accession: F40201  
R:/Claverie, J.M.  
Personal communication, 1992  
A:/Reference number: A40201  
A:/Accession: F40201  
A:/Molecule type: DNA  
A:/Residues: 1-673 <CIA>  
R:/Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:/Title: Identifying coding exons by similarity search: Alu-derived and other potential  
A:/Reference number: A40200; MUID:92241891; PMID:1572661  
A:/Contents: annotation  
C:/Comment: This "Warning" entry is a conceptual translation in all 6 reading frames of c  
in-frame stop codons are shown as 'x'.  
C:/Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:  
Pred. No.: 0 0039 Length: 673  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.34% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x F40201 (1-673)

QY 1839 TGGTGGCTCAGCGCTGTATCCAGCATTGGAG 1874  
Db 4 TrrptleuthrProvalleProhaleutrrpGlu 15

RESULT 7  
178885  
serine/threonine-specific protein kinase (EC 2.7.1.1) STK2 - human  
C:/Species: Homo sapiens (man)  
C:/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 29-Sep-1999  
C:/Accession: 178885  
R:/Lavedakou, E.N.; He, M.; Baptist, E.W.; Craven, R.J.; Cance, W.G.; Welcsh, P.L.; Slamm  
Oncogene 9, 1977-1988, 1994  
A:/Title: Two novel human serine/threonine kinases with homologies to the cell cycle regu  
A:/Reference number: 158396; MUID:94268838; PMID:8208544  
A:/Accession: 178885  
A:/Status: preliminary; translated from GB/EMBL/DBJ  
A:/Molecule type: mRNA  
A:/Residues: 1-841 <RES>  
C:/Cross-references: GB:L20321; NID:9348244; PIDN:AAA36658.1; PID:9348245  
C:/Genetics:  
A:/Gene: GDB:STK2  
A:/Cross-references: GDB:374125  
A:/Map position: 3p21.1-3p21.1  
C:/Superfamily: human serine/threonine-specific protein kinase STK2; protein kinase homol  
C:/Keywords: phosphotransferase  
F:/4-261/Domain: protein kinase homology <KIN>

Alignment Scores:  
Pred. No.: 0 00381 Length: 841  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.34% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880a-7 (1-2818) x 178885 (1-841)

QY 2041 CTCCTGGGTTCAAGCATTTCTCTCCACGCTCC 2006  
Db 478 LeuLeuGlySerSerAspSerProAlaSerAlaSer 489

RESULT 8  
179424  
MHC HLA-DR-beta-1 chain - human (fragment)  
C:/Species: Homo sapiens (man)

C:/Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 23-Jul-1999  
C:/Accession: 179424  
R:/Gregersen, P.K.; Shen, M.; Song, O.  
Proc. Natl. Acad. Sci. U.S.A. 83, 2642-2646, 1986  
A:/Title: Molecular diversity of HLA-DR4 haplotypes.  
A:/Reference number: 159062; MUID:86206008; PMID:3458223  
A:/Accession: 179424  
A:/Status: preliminary; translated from GB/EMBL/DBJ  
A:/Molecule type: mRNA  
A:/Residues: 1-47 <RES>  
C:/Cross-references: GB:M15073; NID:9188370; PIDN:AAA59814.1; PID:9386941  
C:/Superfamily: class II histocompatibility antigen, immunoglobulin homology

Alignment Scores:  
Pred. No.: 0 0548 Length: 47  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.23% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880a-7 (1-2818) x 179424 (1-47)

QY 1847 CACGCTGTATCCAGCATTGGAGGCCGA 1879  
Db 26 HisAlaCysAsnProSerThrLeuGlyGlyArg 36

RESULT 9  
JC5238  
galactosylceramide-1-like protein, GCP - human  
C:/Species: Homo sapiens (man)  
C:/Date: 17-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 07-May-1999  
C:/Accession: JC5238  
R:/Sosa, M.A.G.; De Gasperi, R.; Battistini, S.; Gorman, M.P.; Kolodny, R.; Kolodny, E  
Biochem. Biophys. Res. Commun. 227, 636-641, 1996  
A:/Title: A human kidney cDNA which induces a cell surface protein epitope recognized  
A:/Reference number: JC5238; MUID:97032815; PMID:8878564  
A:/Contents: Kidney  
A:/Accession: JC5238  
A:/Molecule type: mRNA  
A:/Residues: 1-114 <SOS>  
A:/Note: The authors translated the codon GGA for residue 2 as Cys  
C:/Comment: This protein is a cell-surface protein, and belongs to type II transmembra

Alignment Scores:  
Pred. No.: 0 0504 Length: 114  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.23% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880a-7 (1-2818) x JC5238 (1-114)

QY 1912 GACCAAGCTGGCCACATGGCAAAACCCGCTC 1944  
Db 90 AspGlnProGlyGlnHisGlyThrProSer 100

RESULT 10  
B40201  
artificial-warming sequence (translated Alu class B) - human  
C:/Species: Homo sapiens (man)  
C:/Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:/Accession: B40201  
R:/Claverie, J.M.  
Personal communication, 1992  
A:/Reference number: A40201  
A:/Accession: B40201  
A:/Molecule type: DNA  
A:/Residues: 1-301 <CIA>  
R:/Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:/Title: Identifying coding exons by similarity search: Alu-derived and other potenti

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

#### Alignment Scores:

Pred. No.:	0.0459	Length:	301
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.22%	Indels:	0
DB:	4	Gaps:	0

US-09-966-880A-7 (1-2818) x B40201 (1-301)

QY 1790 GGATCTGTATGTCGCCAGCGCTGACCTTGA 1758

Db 266 GlycerCystYrValAlaGlnAlaGlyLeuGlu 276

#### RESULT 11

117239

hypothetical protein DKFZp434B027.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T17239

R:Koehnert, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18722

A:Accession: T17239

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-391 <KOE>

A:Cross-references: EMBL:AL117444

A:Experimental source: adult testis; clone DKFZp434B027

C:Genetics:

A:Note: DKFZp434B027.1

#### Alignment Scores:

Pred. No.:	0.0448	Length:	391
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.22%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-7 (1-2818) x T17239 (1-391)

QY 328 GGAGGTGAACAGGTGACCGGTACGCGGCC 296

Db 276 GlyGlyGluProGlyAspAlaValAlaAla 286

#### RESULT 12

E40201

artifact-warning sequence (translated ALU class E) - human

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000

C:Accession: E40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: E40201

A:Molecule type: DNA

A:Residues: 1-597 <CLA>

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Pred. No.:	0.0425	Length:	673
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.22%	Indels:	0
DB:	4	Gaps:	0

QY 2046 TCCACCTCCTGGSTTCAGGATTCCTCCGCT 2014  
 |||||  
 Db 490 SerThrSerTrpValGlnAlaIleLeuLeuPro 500

Probable receptor-like protein kinase Atg2923300 [Imported] - Arabidopsis thaliana  
N.Alternate names: protein kinase homolog T20D16.7  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 23-Mar-2001  
C.Accession: T00502: A84623  
R:Rounsley, S.D.; Lin, X.; Ketchum, R.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, V. submitted to the EMBL Data Library, November 1997  
A.Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.  
X.Reference number: Z14154

A/Status: translated from GB/EMBL/DDBJ

A; molecule type: DNA

A;Residues: 1-773 &lt;ROU&gt;

A;Cross-references: EMBL:AC002391; NID:g2642427; PTD:g26424333

R. Liliu, X. J. Kaul, S. J. Rounsley, S. D. Shea, T. P. Benito, M. I. Town, C. D. Fujii, C. Y. M., Koo, H., Moffitt, K. S., Cronin, L. A., Shen, M., Yanakiev, S. E., Umayam, L., Tallon, L. E., J. E. Nierman, W. C. J. White, O. Eisen, J. A., Salzberg, S. L., Fraser, C. M., Venter, Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*  
A:Reference number: AB4420; MUID:20083487; PMID:10617497

A;Accession: A84623

A;Status: preliminary

A;Residues: 1-773 &lt;STO&gt;

A/Cross-references: GB:AE002093; NID:g2642433; PIDN:AAB87101.1; GSPDB:GN00139

A;Gene: At2g23300; T20D16.7

A;Map position: 2

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homologs

Pred. No.:	0.0419	length:	773
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.22%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-7 (1-2818) x T00502 (1-773)

2816 TTTTTCAGTTTATTC 2784

Db 15 PhePhePhePhePheSerSerLeuphe 25

```
Search completed: July 7, 2003, 23:38:59
Job time : 69.5 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 7, 2003, 22:17:42 ; Search time 32.5 Seconds

(without alignments)  
7192.633 Million cell updates/sec

Title: US-09-966-880A-7

Perfect score: 897  
Sequence: 1 agagaccatcattatga.....aaaaaaaaaaaaaaaaaa 2818

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Xgapop 60.0 , Xgapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xip  
-O=/cgn2.1/USPRO/US09966880/runat\_07072003\_142233\_22064/app\_query.fasta\_1.3015  
-DB=SwissProt\_40 -QFMT=fastan -SUFFIX=oligo -rnp -MINMATCH=0.1 -LOOP=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR.SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09966880 -CGN 1.1.31 -runat\_07072003\_142233\_22064 -MCP=6 -ICU=3  
-NO\_MMAP -LARGESOURCY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONCLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match Length	DB ID	Description
C 1	23	2.6	585 1	ALU5_HUMAN
C 2	23	2.6	591 1	ALU8_HUMAN
C 3	23	2.6	593 1	ALU7_HUMAN
C 4	22	2.4	593 1	ALU6_HUMAN
C 5	21	2.3	585 1	ALU8_HUMAN
C 6	21	2.3	591 1	ALU8_HUMAN
C 7	21	2.3	591 1	ALU7_HUMAN
C 8	20	2.2	587 1	ALU2_HUMAN
C 9	20	2.2	587 1	ALU2_HUMAN
C 10	20	2.2	593 1	ALU6_HUMAN
C 11	18	2.0	591 1	ALU6_HUMAN
C 12	18	2.0	603 1	ALU4_HUMAN
C 13	13	1.4	587 1	ALU2_HUMAN
C 14	13	1.4	587 1	ALU3_HUMAN
C 15	13	1.4	591 1	ALU1_HUMAN
C 16	13	1.4	603 1	ALU4_HUMAN
C 17	12	1.3	841 1	NEK4_HUMAN
C 18	12	1.3	881 1	PKP2_HUMAN

C 19	11	1.2	629 1	2195_HUMAN	014628 homo sapien
C 20	10	1.1	807 1	AFSK_STRGR	P34742 streptomyc
C 21	9	1.0	167 1	YK20_YEAST	P36133 saccharomyc
C 22	9	1.0	285 1	IFZB_YEAST	P09064 saccharomyc
C 23	9	1.0	386 1	LE11_SULFO	O974x3 sulfobus
C 24	9	1.0	670 1	SR72_CANFA	P33731 canis famli
C 25	9	1.0	670 1	SR72_HUMAN	O76094 homo sapien
C 26	9	1.0	807 1	AFSK_STRGR	P34742 streptomyc
C 27	9	1.0	843 1	BLVR_BOVIN	O03368 bos taurus
C 28	9	1.0	1888 1	YD27_SCHPO	O14207 schizosach
C 29	8	0.9	91 1	YMBG_VACCC	P20547 vacocia vi
C 30	8	0.9	105 1	YMBG_VACCC	P15611 paramecium
C 31	8	0.9	111 1	YG2C_YEAST	P33245 saccharomyc
C 32	8	0.9	124 1	YNY3_HUMAN	P20931 homo sapien
C 33	8	0.9	130 1	YNO3_YEAST	P33633 saccharomyc
C 34	8	0.9	139 1	YED3_YEAST	P32633 saccharomyc
C 35	8	0.9	144 1	YHP5_YEAST	P33808 saccharomyc
C 36	8	0.9	145 1	YNR9_YEAST	P53880 saccharomyc
C 37	8	0.9	174 1	NEUO_YEAST	O94x8 mus musculu
C 38	8	0.9	174 1	NEUO_MOUSE	P12760 rattus norv
C 39	8	0.9	179 1	VC59_HSVB	P28983 equine herp
C 40	8	0.9	202 1	LIF_BOVIN	O27956 bos taurus
C 41	8	0.9	202 1	LIF_MUSVI	O62728 mustela vis
C 42	8	0.9	217 1	KSL_HYDAT	P38978 hydra atten
C 43	8	0.9	224 1	Y364_MYCGE	P47604 mycoplasma
C 44	8	0.9	245 1	TMAB_HUMAN	O9ukr8 homo sapien
C 45	8	0.9	253 1	ASP_PLAFS	P13825 plasmodium

# ALIGNMENTS

RESULT 1	ALU5_HUMAN	STANDARD:	PRT:	585 AA.
ID	ALU5_HUMAN			
AC	P39192;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DE	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Alu subfamily SC sequence contamination warning entry.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=95021759; PubMed=7935834;			
RA	Claverie J.-M.; Makalowski W.;			
RT	"Alu alert.";			
RL	Nature 371:752-752(1994).			
RN	[2]			
RP	CONCEPT.			
RP	MEDLINE=92241891; PubMed=1572661;			
RA	Claverie J.-M.;			
RT	"Identifying coding exons by similarity search: alu-derived and other			
RT	potentially misleading protein sequences.";			
RL	Genomics 12:838-841(1992).			
RN	[3]			
RP	ALU FAMILIES CLASSIFICATION.			
RP	MEDLINE=8833009; PubMed=3138422;			
RA	Quantin Y.;			
RT	"The Alu family developed through successive waves of fixation			
RT	closely connected with primate lineage history.";			
RL	J. Mol. Evol. 27:194-202(1988).			
RN	[4]			
RP	ALU FAMILIES CLASSIFICATION.			
RP	MEDLINE=91178815; PubMed=1706781;			
RA	Jurka J.; Milosavljevic A.;			
RT	"Reconstruction and analysis of human Alu genes.";			
RL	J. Mol. Evol. 32:105-121(1991).			
CC	-I- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE			
CC	THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING			
CC	CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX			
CC	FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU			

```

CC REPEATS.
CC
CC MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'xxx' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; U14571; ?; NOT_ANNOTATED_CDS.
CC
CC FT Hypothetical protein.
CC FT DOMAIN 1 95 FRAME-1.
CC FT DOMAIN 99 193 FRAME-2.
CC FT DOMAIN 197 291 FRAME-3.
CC FT DOMAIN 295 389 FRAME-4.
CC FT DOMAIN 393 487 FRAME-5.
CC FT DOMAIN 491 585 FRAME-6.
CC SQ SEQUENCE 585 AA; 63957 MW; 4AE8BC4F93650A7 CNC64;
CC
CC
CC Alignment Scores:
CC Pred. NO.: 1.38e-14 Length: 585
CC Score: 23.00 Matches: 23
CC Percent Similarity: 100.00% Conservative: 0
CC Best Local Similarity: 100.00% Mismatches: 0
CC Query Match: 2.56% Indels: 0
CC DB: 1 Gaps: 0
CC
CC US-09-966-880A-7 (1-2818) x ALU5_HUMAN (1-585)
CC
CC QY 2054 CTGCAACCTCACCCTCGGSGTGAAGGAGATTCTCGCTGACCTGCCAAGTAGCTGG 1995
CC |||||||
CC Db 414 LeuGlnProProProProGlyPheIysArgPheSerCysLeuSerLeuProSerSerTrp 433
CC
CC QY 1994 GATTAACAAG 1986
CC |||||||
CC Db 434 AspIlyArg 436
CC
CC RESULT 2
CC ALU8_HUMAN
CC ID ALU8_HUMAN STANDARD; PRT; 591 AA.
CC AC P39195.
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last annotation update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Alu subfamily SX sequence contamination warning entry.
CC OS Homo sapiens (Human).
CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC NCBI_taxid=9606;

```

RP [1] SEQUENCE FROM N.A.  
 RX MEDLINE=95021756; PubMed=7935634;  
 RA Claverie U.-M., Makalowski W.;  
 RT "Alu alert."  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie U.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences."  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history."  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Miosavljivic A.;  
 RT "Reconstruction and analysis of human Alu genes."  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: U14574; -; NOT\_ANNOTATED\_CDS.  
 CC KW Hypothetical protein.  
 CC FT DOMAIN 1  
 CC FT 100 195 FRAME-1.  
 CC FT 199 294 FRAME-2.  
 CC FT DOMAIN 298 393 FRAME-3.  
 CC FT 397 492 FRAME-4.  
 CC FT DOMAIN 496 591 FRAME-5.  
 CC FT 591 64395 MM; AC081540BA6B280 CRC64; FRAME-6  
 CC SEQUENCE





```

RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentín Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history."
RL J. Mol. Evol. 27:194-202(1986).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA J. M. Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14572; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1
FT DOMAIN 101 196 FRAME-1.
FT DOMAIN 200 295 FRAME-2.
FT DOMAIN 299 395 FRAME-3.
FT DOMAIN 399 494 FRAME-4.
FT DOMAIN 593 593 FRAME-5.
FT DOMAIN 593 593 FRAME-6.
SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.: 1 45e-13 Length: 593
Score: 22.00 Matches: 22
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 1 Gaps: 0

US-09-966-880a-7 (1-2818) x ALU6_HUMAN (1-593)
QY 2051 CAACCTCAACCTCTGGGTTCAAGCATCTCTGCTCAGGCTCCAGTAGTGGGAT 1992
Db 322 GlnProProProProProlylPhelysArgPheSerCysIleuSerLeuProSerSerTrpPasp 341

```

```

OY 1991 TACAGG 1986
DB 342 TyrArg 343

RESULT 5
ALU5_HUMAN
ID ALU5_HUMAN STANDARD: PRT: 585 AA.
AC P39192:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makolowski W.;
RT "Alu alert."
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572651;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences."
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentín Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history."
RL J. Mol. Evol. 27:194-202(1986).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA J. M. Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U14571; -; NOT\_ANNOTATED\_CDS.  
KW Hypothetical protein  
FT DOMAIN 1 95 FRAME-1.  
FT DOMAIN 99 193 FRAME-2.  
FT DOMAIN 197 291 FRAME-3.  
FT DOMAIN 295 389 FRAME-4.  
FT DOMAIN 393 487 FRAME-5.  
FT DOMAIN 491 585 FRAME-6.  
SQ SEQUENCE 585 AA; 63957 MW; 46EB8C4F493650A7 CRC64;  
  
Alignment Scores:  
Pred. No.: 1.53e-12 Length: 585  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.34% Indels: 0  
DB: Gaps: 0  
US-09-966-880a-7 (1-2818) x ALU5\_HUMAN (1-585)  
QY 1993 TCCGACCTACTTGGAGCGTGGAGCAGAGAAATCGCTTGACCCAGAGAGTGAGTTGC 2052  
DB 250 SerGlnLeuLeuGlyArgLeuArgGlnGlnuAsnArgLeuAsnProGlyGlyGlyGlyCys 269  
QY 2053 AGT 2055  
DB 270 Ser 270  
  
RESULT 6  
ALU8\_HUMAN  
ID ALU8\_HUMAN STANDARD; PRT; 591 AA.  
AC P39195;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Alu subfamily SX sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M.; Makalowski W.;  
RT "Alu alert."  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences."  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history."  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Turka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes."  
RL J. Mol. Evol. 32:105-121(1991).

CC MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC COON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U14574; -; NOT\_ANNOTATED\_CDS.  
KW Hypothetical protein.  
FT DOMAIN 1 96 FRAME-1.  
FT DOMAIN 100 195 FRAME-2.  
FT DOMAIN 199 294 FRAME-3.  
FT DOMAIN 298 393 FRAME-4.  
FT DOMAIN 397 492 FRAME-5.  
FT DOMAIN 496 591 FRAME-6.  
SQ SEQUENCE 591 AA; 64395 MW; AC8154A08A6B280 CRC64;  
  
Alignment Scores:  
Pred. No.: 1.53e-12 Length: 591  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.34% Indels: 0  
DB: Gaps: 0  
US-09-966-880a-7 (1-2818) x ALU8\_HUMAN (1-591)  
QY 1993 TCCGACCTACTTGGAGCGTGGAGCAGAGAAATCGCTTGACCCAGAGAGTGAGTTGC 2052  
DB 154 SerGlnLeuLeuGlyArgLeuArgGlnGlnuAsnArgLeuAsnProGlyGlyGlyGlyCys 173  
QY 2053 AGT 2055  
DB 174 Ser 174  
  
RESULT 7  
ALU7\_HUMAN  
ID ALU7\_HUMAN STANDARD; PRT; 593 AA.  
AC P39194;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily SQ sequence contamination warning entry.

OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M.; Makalowski W.;  
 RT "Alu alert";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J.; Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN AN ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U14573; ; NOT\_ANNOTATED\_CDS.  
 KW Hypothetical protein.  
 FT DOMAIN 1 97 FRAME-1.  
 FT DOMAIN 101 196 FRAME-2.  
 FT DOMAIN 200 295 FRAME-3.  
 FT DOMAIN 299 395 FRAME-4.

FT DOMAIN 399 494 FRAME-5.  
 FT DOMAIN 498 593 FRAME-6.  
 SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;  
 Alignment Scores:  
 Pred. No.: 1.53e-12 Length: 593  
 Score: 21.00 Matches: 21  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.348 Indels: 0  
 DB: 1 Caps: 0  
 US-09-966-880a-7 (1-2818) x ALU7\_HUMAN (1-593)  
 QY 1993 TCCGACGCTACTGGAGGCGTCTGAGCAGAGATCGCTTGACCCAGAGGTGAGGTTGC 2052  
 DB 155 SerGlnLeuLeuGlyArgLeuArgGlnGlnAsnArgLeuAsnProGlyGlyGlyCys 174  
 QY 2053 AGT 2055  
 DB 175 Ser 175  
 RESULT 8  
 ALU2\_HUMAN STANDARD; PRT; 587 AA.  
 ID ALU2\_HUMAN  
 AC P39189;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily 5B sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M.; Makalowski W.;  
 RT "Alu alert";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J.; Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN

CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: U14568; -; NOT\_ANNOTATED\_CDS.  
 CC KM Hypothetical protein.  
 CC FT DOMAIN 1 96 FRAME-1.  
 CC FT DOMAIN 100 194 FRAME-2.  
 CC FT DOMAIN 198 292 FRAME-3.  
 CC FT DOMAIN 296 391 FRAME-4.  
 CC FT DOMAIN 395 489 FRAME-5.  
 CC FT DOMAIN 493 587 FRAME-6.  
 CC SQ SEQUENCE 587 AA; 63703 MW; 3EAB3E3E3929203 CRC64;  
 CC  
 CC Alignment Scores:  
 CC Pred. No.: 1.61e-11 Length: 587  
 CC Score: 20.00 Matches: 20  
 CC Percent Similarity: 100.00% Conservative: 0  
 CC Best Local Similarity: 100.00% Mismatches: 0  
 CC Query Match: 2.23% Indels: 0  
 CC DB: 1 Gaps: 0  
 CC  
 CC US-09-966-880a-7 (1-2818) x ALU2\_HUMAN (1-587)  
 CC  
 CC QY 1889 TCCGCCCGGCTCGGCTCCCAAGTGTGAGATTACAGGCGGTGACCCACCGCCGCC 1830  
 CC |||||||  
 CC Db 372 SerAlaArgLeuGlyLeuProLysCysTyrAspTyrAlaArgGluProArgProAla 391  
 CC  
 CC RESULT 9  
 CC ALU3\_HUMAN  
 CC ID ALU3\_HUMAN STANDARD; PRT; 587 AA.  
 CC AC P39190;  
 CC DT 01-FEB-1995 (Rel. 31, Created)  
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE ALU subfamily SBI sequence contamination warning entry.  
 CC OS Homo sapiens (Human).  
 CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC CC NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=95021758; PubMed=7935834;  
 CC RA Claverie J.-M.; Makalowski W.;  
 CC RT "Alu alert."  
 CC RL Nature 371:752-752(1994).  
 CC RP [2]  
 CC RP CONCEPT.  
 CC RX MEDLINE=92241891; PubMed=1572661;  
 CC RA Claverie J.-M.;  
 CC RT "Identifying coding exons by similarity search: alu-derived and other  
 CC RT potentially misleading protein sequences."  
 CC RL Genomics 12:838-841(1992).

RN [3]  
 RP ALU FAMILIES CLASSIFICATION  
 RX MEDLINE=88333009; PubMed=318422;  
 RA Quenian Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history."  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Miosavljewic A.;  
 RT "Reconstruction and analysis of human Alu genes."  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'xxx' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: U14569; -; NOT\_ANNOTATED\_CDS.  
 CC KM Hypothetical protein.  
 CC FT DOMAIN 1 96 FRAME-1.  
 CC FT DOMAIN 100 194 FRAME-2.  
 CC FT DOMAIN 198 292 FRAME-3.  
 CC FT DOMAIN 296 391 FRAME-4.  
 CC FT DOMAIN 395 489 FRAME-5.  
 CC FT DOMAIN 493 587 FRAME-6.  
 CC SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEP235 CRC64;  
 CC  
 CC Alignment Scores:  
 CC Pred. No.: 1.61e-11 Length: 587  
 CC Score: 20.00 Matches: 20  
 CC Percent Similarity: 100.00% Conservative: 0  
 CC Best Local Similarity: 100.00% Mismatches: 0  
 CC Query Match: 2.23% Indels: 0  
 CC DB: 1 Gaps: 0  
 CC  
 CC US-09-966-880a-7 (1-2818) x ALU3\_HUMAN (1-587)  
 CC  
 CC QY 1889 TCCGCCCGGCTCGGCTCCCAAGTGTGAGATTACAGGCGGTGACCCACCGCCGCC 1830  
 CC |||||||  
 CC Db 372 SerAlaArgLeuGlyLeuProLysCysTyrAspTyrAlaArgGluProArgProAla 391

```

10 RESULT 10
11 ALU5_HUMAN
12 ID ALU5_HUMAN STANDARD: PRT; 593 AA.
13 AC P39193;
14 DT 01-FEB-1995 (Rel. 31, Created)
15 DT 01-FEB-1995 (Rel. 31, Last sequence update)
16 DT 16-OCT-2001 (Rel. 40, Last annotation update)
17 DE ALU subfamily SP sequence contamination warning entry.
18 OS Homo sapiens (Human).
19 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
20 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
21 OX NCBI_Taxid=9606;
22 RN [1]
23 RP SEQUENCE FROM N. A.
24 RX MEDLINE=95021758; PubMed=7935834;
25 RA Claverie J.-M., Makalowski W.;
26 RT "Alu alert.";
27 RL Nature 371:752-752(1994).
28 RN [2]
29 RP CONCEPT.
30 RX MEDLINE=92241891; PubMed=1572661;
31 RA Claverie J.-M.;
32 RT "Identifying coding exons by similarity search: alu-derived and other
33 RT potentially misleading protein sequences.";
34 RL Genomics 12:838-841(1992).
35 RN [3]
36 RP ALU FAMILIES CLASSIFICATION,
37 MEDLINE=88333009; PubMed=3138422;
38 RA Quentin Y.;
39 RT "The Alu family developed through successive waves of fixation
40 RT closely connected with primate lineage history.";
41 RL J. Mol. Evol. 27:194-202(1988).
42 RN [4]
43 RP ALU FAMILIES CLASSIFICATION,
44 MEDLINE=91178815; PubMed=1706781;
45 RA Jurka J., Milosavljevic A.;
46 RT "Reconstruction and analysis of human Alu genes.";
47 RL J. Mol. Evol. 32:105-121(1991).
48 RN [5]
49 RP MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
50 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
51 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
52 FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
53 REPEATS.
54 RN [6]
55 RP MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
56 COON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
57 RN [7]
58 RP CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
59 POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
60 ACID SEQUENCES.
61 RN [8]
62 RP CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
63 PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
64 ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
65 ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
66 CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CNAS
67 LIGATED WITH ALU-DERIVED SEQUENCE IN AN ORIENTATIONAL, ALTHOUGH ALU
68 ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
69 GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
70 CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
71 OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
72 CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
73 DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
74 THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
75 BEING REPORTED.
76 RN [9]
77 RP CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
78 WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
79 PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
80 CODING NUCLEOTIDE SEQUENCE.
81
82 -----
83 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
84 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
85 CC the European Bioinformatics Institute. There are no restrictions on its
86 CC use by non-profit institutions as long as its content is in no way
87 CC modified and this statement is not removed. Usage by and for commercial

```

```
CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL; U14572; -; NOT_ANNOTATED_CDS.  
KM Hypothetical protein.  
FT DOMAIN 1 97 FRAME-1.  
FT DOMAIN 101 196 FRAME-2.  
FT DOMAIN 200 295 FRAME-3.  
FT DOMAIN 299 395 FRAME-4.  
FT DOMAIN 399 494 FRAME-5.  
FT DOMAIN 498 593 FRAME-6.  
SQ SEQUENCE 593 AA; 64603 MW; 136EB344AACD12A2 CRC64;  
  
Alignment Scores:  
Pred. No.: 1.61e-11 Length: 593  
Score: 20.00 Matches: 20  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.23% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-966-880A-7 (1-2818) x ALU6_HUMAN (1-593)  
CY 1993 TCCACGCTACTTGGCAGCGCTGAGGCAGAGAATTCGTGAACCACGAGAGTGAGCTGC 2052  
Db 155 SerpinleuenglyArgLeuArgIngluAsnArgLeuasnProglylglygycys 174  
  
RESULT 11  
ALU1_HUMAN  
ID ALU1_HUMAN STANDARD; PRT; 591 AA.  
AC P39188;  
DT 01-FEB-1995 (Rel. 31, Created)  
DI 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily J sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI_Taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935634;  
RA Claverie J.-M.; Makalowski W.;  
RT "Alu alert."  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=8833509; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RL closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -1 MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -1 MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON. 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -1 CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
```

```

CC POLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIVATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14567; ; NOT_ANNOTATED_CDS.
CC
CC DR Hypothetical protein.
CC KM
CC FT DOMAIN 1 96 FRAME-1.
CC FT DOMAIN 100 195 FRAME-2.
CC FT DOMAIN 199 294 FRAME-3.
CC FT DOMAIN 298 393 FRAME-4.
CC FT DOMAIN 397 492 FRAME-5.
CC FT DOMAIN 496 591 FRAME-6.
CC SQ SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;

Alignment Scores:
Pred. No.: 1.78e-09 Length: 591
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.00% Indels: 0
DB: Gaps: 1

US-09-966-880a-7 (1-2818) x ALU1_HUMAN (1-591)
QY 1883 CGCGCTGGCGCTCCCAAGTGGTGGGATTACAGGCGTGACGCCACACGCCCGCCG 1830
DB 574 ArgLeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArgProAla 591

RESULT 12
ALU4_HUMAN
ID ALU4_HUMAN STANDARD; PRT; 603 AA.
AC P39191;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SB2 sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.

```

```

RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J.; Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14570; ; NOT_ANNOTATED_CDS.
CC
CC DR Hypothetical protein.
CC KM
CC FT DOMAIN 1 98 FRAME-1.
CC FT DOMAIN 102 199 FRAME-2.
CC FT DOMAIN 202 300 FRAME-3.
CC FT DOMAIN 304 401 FRAME-4.
CC FT DOMAIN 405 502 FRAME-5.
CC FT DOMAIN 506 603 FRAME-6.
CC SQ SEQUENCE 603 AA; 65272 MW; B8AAD0AD46BEA114 CRC64;

Alignment Scores:
Pred. No.: 1.77e-09 Length: 603
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.00% Indels: 0
DB: Gaps: 1

```

US-09-966-880a-7 (1-2818) x ALU4\_HUMAN (1-603)

QY 1883 CGGCTCGGCTCCCAAGTGTGATTCAGCGTGCAGCCAGCCGCGGC 1830  
 |||||||  
 Db 586 ArgLengGlyLeuProLysCysTrpAspTrpArgGluProProArgProAla 603

RESULT 13

ALU2\_HUMAN STANDARD; PRT; 587 AA.

AC P39189;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily SB sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 closely connected with primate lineage history";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC - MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 REPEATS.  
 CC - MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC - CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 ACID SEQUENCES.  
 CC - CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN  
 ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 ELEMENTS (SPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 BEING REPORTED.  
 CC - CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE  
 CODING NUCLEOTIDE SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: U14568; NOT\_ANNOTATED\_CDS.  
 DR  
 KM Hypothetical protein.  
 FT DOMAIN 1 96 FRAME-1.  
 FT DOMAIN 100 194 FRAME-2.  
 FT DOMAIN 198 292 FRAME-3.  
 FT DOMAIN 296 391 FRAME-4.  
 FT DOMAIN 395 489 FRAME-5.  
 FT DOMAIN 493 587 FRAME-6.  
 SQ SEQUENCE 587 AA; 63703 MW; 3EAB3EE3E3929203 CRC64;

Alignment Scores:

Pred. No.:	0.00023	Length:	587
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.45%	Indels:	0
DB:	1	Gaps:	0

US-09-966-880a-7 (1-2818) x ALU2\_HUMAN (1-587)

QY 1842 TGCTCAGCGCTGTATCCAGCACTTGGAGGCGCGAG 1880  
 |||||||  
 Db 5 TripleThrProValIleProAlaLeuTrpGluIleAlu 17

RESULT 14

ALU3\_HUMAN STANDARD; PRT; 587 AA.

AC P39190;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily SBI sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 closely connected with primate lineage history";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC - MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX

CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1 MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON. 'xxx' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1 CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1 CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1 CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U14567; -; NOT\_ANNOTATED\_CDS.  
 CC KW Hypothetical protein.  
 CC FT DOMAIN 1 96 FRAME-1.  
 CC FT DOMAIN 100 194 FRAME-2.  
 CC FT DOMAIN 198 292 FRAME-3.  
 CC FT DOMAIN 296 391 FRAME-4.  
 CC FT DOMAIN 395 489 FRAME-5.  
 CC FT DOMAIN 493 587 FRAME-6.  
 CC SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEP235 CRC64;  
 CC  
 CC Alignment Scores:  
 CC Pred. No.: 0.00023 Length: 587  
 CC Score: 13.00 Matches: 13  
 CC Percent Similarity: 100.00% Conservative: 0  
 CC Best Local Similarity: 100.00% Mismatches: 0  
 CC Query Match: 1.45% Indels: 0  
 CC DB: 1 Gaps: 0  
 CC  
 CC US-09-966-880a-7 (1-2818) x ALU3\_HUMAN (1-587)  
 CC QY 1842 TGGGTACGCGCTGTATCCAGACACTTTGGAGAGCCGAG 1880  
 CC Db 5 TrrpleuThrProValIleProAlaIleuTrpGluAlaGlu 17  
 CC  
 CC RESULT 15  
 CC ALU1\_HUMAN  
 CC ID ALU1\_HUMAN STANDARD; PRT; 591 AA.  
 CC AC P39186;  
 CC DT 01-FEB-1995 (Rel. 31, Created)  
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Alu subfamily J sequence contamination warning entry.  
 CC OS Homo sapiens (Human).  
 CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=95021758; PubMed=7935834;

RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert."  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences."  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history."  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes."  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1 MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1 MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON. 'xxx' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1 CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1 CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1 CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U14567; -; NOT\_ANNOTATED\_CDS.  
 CC KW Hypothetical protein.  
 CC FT DOMAIN 1 96 FRAME-1.  
 CC FT DOMAIN 100 195 FRAME-2.  
 CC FT DOMAIN 198 294 FRAME-3.  
 CC FT DOMAIN 298 393 FRAME-4.  
 CC FT DOMAIN 397 492 FRAME-5.  
 CC FT DOMAIN 496 591 FRAME-6.  
 CC SQ SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;  
 CC  
 CC Alignment Scores:  
 CC Pred. No.: 0.000229 Length: 591  
 CC Score: 13.00 Matches: 13



Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.45% Indels: 0  
 DB: 1 Gaps: 0

US-09-966-880a-7 (1-2818) x ALU1\_HUMAN (1-591)

QY 1842 TGGCTCAGCGCTGTATCCAGCAGCTTGGAGCCGAG 1880  
 ||||||||||||||||||||||||||||||||||||  
 Db 5 TrrpleuthrProValilleProAlaLeutrPGluAlaGlu 17

Search completed: July 7, 2003, 23:33:14  
 Job time : 46.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 7, 2003, 23:16:57 ; Search time 103.5 Seconds  
(without alignments)  
11220.111 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 897

Sequence: 1 agagaacatcataatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09966880/runat\_07072003\_142233\_22077/app\_query.fasta.1.3015  
-DB=SPREMBL\_21 -QFMT=fastan -SUFFIX=olig.rspct -MINMATCH=0.1 -LOCPCL=0  
-LISTEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-OUTFMT=pro -NORExt -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US09966880.ecgn\_1\_1\_155\_@runat\_07072003\_142233\_22077 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

SPREMBL\_21:.\*  
1: sp\_archaea:.\*  
2: sp\_bacteria:.\*  
3: sp\_fungi:.\*  
4: sp\_human:.\*  
5: sp\_invertebrate:.\*  
6: sp\_mammal:.\*  
7: sp\_mhc:.\*  
8: sp\_organelle:.\*  
9: sp\_phage:.\*  
10: sp\_plant:.\*  
11: sp\_rodent:.\*  
12: sp\_virus:.\*  
13: sp\_vertebrate:.\*  
14: sp\_unclassified:.\*  
15: sp\_rvivirus:.\*  
16: sp\_bacteriophage:.\*  
17: sp\_archaeophages:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	198	22.1	198	4	Q9GZX7

2	42	4.7	198	11	Q9WVE0	Q9WVE0 mus musculi
3	18	2.0	72	4	Q8TB48	Q8TB48 homo sapien
4	18	2.0	368	4	Q15662	Q15662 homo sapien
5	17	1.9	133	4	Q9H8X0	Q9H8X0 homo sapien
6	17	1.9	318	4	Q96QJ7	Q96QJ7 homo sapien
7	17	1.9	423	4	Q9NMF0	Q9NMF0 homo sapien
8	16	1.8	155	4	Q9H800	Q9H800 homo sapien
9	16	1.8	181	4	Q9NX13	Q9NX13 homo sapien
10	16	1.8	211	4	Q8TDM0	Q8TDM0 homo sapien
11	16	1.8	231	4	Q9NR08	Q9NR08 homo sapien
12	16	1.8	232	4	Q9H5R3	Q9H5R3 homo sapien
13	16	1.8	375	4	Q60448	Q60448 homo sapien
14	15	1.7	126	4	Q9UHT2	Q9UHT2 homo sapien
15	15	1.7	166	4	Q96LU6	Q96LU6 homo sapien
16	15	1.7	1596	4	Q9HCL6	Q9HCL6 homo sapien
17	14	1.6	76	4	Q9P191	Q9P191 homo sapien
18	14	1.6	102	4	Q8WY68	Q8WY68 homo sapien
19	14	1.6	118	4	Q9H387	Q9H387 homo sapien
20	14	1.6	449	4	Q9NX12	Q9NX12 homo sapien
21	13	1.4	50	4	Q96EM6	Q96EM6 homo sapien
22	13	1.4	90	4	Q96TGI	Q96TGI homo sapien
23	13	1.4	102	4	Q9P147	Q9P147 homo sapien
24	13	1.4	130	4	Q9HBS7	Q9HBS7 homo sapien
25	13	1.4	133	4	Q96JRS	Q96JRS homo sapien
26	13	1.4	174	4	Q9H926	Q9H926 homo sapien
27	13	1.4	294	4	Q96CC2	Q96CC2 homo sapien
28	13	1.4	296	4	Q8WWD5	Q8WWD5 homo sapien
29	13	1.4	296	7	Q29890	Q29890 homo sapien
30	13	1.4	328	4	Q96BG8	Q96BG8 homo sapien
31	13	1.4	382	4	Q96AL7	Q96AL7 homo sapien
32	13	1.4	383	4	Q8TEJ6	Q8TEJ6 homo sapien
33	13	1.4	390	4	Q9HAL3	Q9HAL3 homo sapien
34	13	1.4	666	4	P78525	P78525 homo sapien
35	13	1.4	694	4	Q96JTS	Q96JTS homo sapien
36	13	1.4	1406	4	Q15082	Q15082 homo sapien
37	12	1.3	26	4	Q9BRJ7	Q9BRJ7 homo sapien
38	12	1.3	119	4	Q8WY65	Q8WY65 homo sapien
39	12	1.3	128	4	Q8WYX2	Q8WYX2 homo sapien
40	12	1.3	129	4	Q9H9H0	Q9H9H0 homo sapien
41	12	1.3	141	4	Q9H6Y8	Q9H6Y8 homo sapien
42	12	1.3	169	4	Q9H397	Q9H397 homo sapien
43	12	1.3	827	4	Q9P225	Q9P225 homo sapien
44	12	1.3	827	4	Q9P1C3	Q9P1C3 homo sapien
45	11	1.2	46	4	Q9P1C3	Q9P1C3 homo sapien

## ALIGNMENTS

### RESULT 1

ID	Q9GZX7	PRELIMINARY	PRT	198 AA.
AC	Q9GZX7			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Activation-induced cytidine deaminase.			
GN	AID.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20408890; PubMed=10950930;			
RA	Muto T., Muramatsu M., Tanikawa K., Honjo T.;			
RT	"Isolation, tissue distribution and chromosomal localization of the			
RT	human activation-induced cytidine deaminase (AID) gene.";			
RL	Genomics 68:85-88(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20460541; PubMed=11007475;			
RA	Rey P., Muto T., Levy Y., Geissmann F., Pieban A., Sanal O.,			
RA	Catalan N., Forveille M., Dufourcq-Lagelouse R., Gennery A.,			



US-09-966-880a-7 (1-2818) x Q8TB48 (1-72)

QY 2057 TTACTGCAACCTCCAGCTCTGGGTTTCAGCGATTCTCTGCTCAGCTCCCA 2004  
 DB 18 LeuLeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuPro 35

# RESULT 4

ID Q15662 PRELIMINARY; PRT; 368 AA.  
 AC Q15662;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)  
 DE Transformation-related protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EPITHELIUM;  
 RA Shen H., Steinberg M.L.;  
 RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: L24521; AAA36776.1; -;  
 FT NON\_TER 1  
 SQ SEQUENCE 368 AA; 42029 MW; A8B79E59E3BA2B0 CRC64;

Alignment Scores:  
 Pred. No.: 1.83e-09 Length: 368  
 Score: 18.00 Matches: 18  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.00% Indels: 0  
 DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x Q15662 (1-368)

QY 2039 CTTGGGTCAGCGATTCTCTGCTCAGCTCCAGTTCAGGATTACAGG 1986  
 DB 25 ProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArg 42

# RESULT 5

ID Q9H8K0 PRELIMINARY; PRT; 133 AA.  
 AC Q9H8K0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE CDNA FLJ13520 fis, clone PLACE1005828.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.,  
 RA "NEO human cDNA sequencing project."  
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK033582; BAB14616.1; -;  
 SQ SEQUENCE 133 AA; 14646 MW; F74020BA4B9CCF3E CRC64;

Alignment Scores:  
 Pred. No.: 2.52e-08 Length: 133  
 Score: 17.00 Matches: 17  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.89% Indels: 0  
 DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x Q9H8K0 (1-133)

QY 2441 CTTCTGGGTCAGCGATTCTCTGCTCAGCTCCCAAGTACTGGGATT 1991  
 DB 92 LeuLeuLysSerSerAspSerProValSerAlaSerGlnValAlaGlyIle 108

# RESULT 6

ID Q96QL7 PRELIMINARY; PRT; 318 AA.  
 AC Q96QL7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Similar to CAMP responsive element binding protein-like 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PROSTATE;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC008394; AA08394.1; -;  
 SQ SEQUENCE 318 AA; 33894 MW; D70E20D5409C165B CRC64;

Alignment Scores:  
 Pred. No.: 2.2e-08 Length: 318  
 Score: 17.00 Matches: 17  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.90% Indels: 0  
 DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x Q96QL7 (1-318)

QY 1830 GGGGGGGCGGCGGCTCAGCGCTGTAATCCAGCACTTGGAGGCCGAG 1880  
 DB 285 GAlYArgAlaTrpTrpLeuThrProValIleProAlaLeuTrpGlnAlaGln 301

# RESULT 7

ID Q9NWF0 PRELIMINARY; PRT; 423 AA.  
 AC Q9NWF0;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 46.5 kDa protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanaka T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,  
 RA "NEO human cDNA sequencing project."  
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RL -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.  
 DR EMBL: AK000928; BA91431.1; -;

DR InterPro: IPR001630; LeuZIP\_CREB.  
 DR InterPro: IPR004827; TF\_BZIP.  
 DR Pfam: PF00170; bzip.1.  
 DR PRINTS: PR00041; LEUZIP\_CREB.  
 DR SMART: SM00338; BRLZ.1.  
 DR PROSITE: PS00036; BZIP\_BASIC.1.  
 KW DNA-binding; Nuclear protein.  
 SQ SEQUENCE 423 AA; 46463 MW; 348A5EFC5A1E977 CRC64;

## Alignment Scores:

Pred. No.: 2.11e-08 Length: 423  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.90% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x Q9NMF0 (1-423)

QY 1830 GGGCGGGCGTGGTGGCTCAGCCCTGTAATCCAGACGCTTGGAGGCCGAG 1880

DB 390 GtYrGAlatPrpPneutPrProValIleProAlaLeutPrGluAlaGlu 406

## RESULT 8

Q9H800 PRELIMINARY; PRT; 155 AA.

AC Q9H800; 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DE Hypothetical 17.1 kDa protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EMBRIO;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Wagaatsuma M., Hosolri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,

RA Niinomiya K., Iwayanagi T.;

RT "NEDO human cDNA sequencing project."

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF024093; BAB14823.1; -

SQ SEQUENCE 155 AA; 17101 MW; D11F9640C176E7F6 CRC64;

## Alignment Scores:

Pred. No.: 2.9e-07 Length: 155  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.78% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x Q9H800 (1-155)

QY 2032 TCAGGAGATTCTGCTGCTCAGCCCTCCCAAGTAGCTGGATTACAGGT 1985

DB 67 SerSerAspSerProAlaSerAlaSerGlnValAlaGlyIleThrGly 82

## RESULT 9

Q9NX13 PRELIMINARY; PRT; 181 AA.

AC Q9NX13; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE CDNA FLJ20495 f1s, clone KAT08572.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EMBRIO;

RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,

RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,

RA Nakamura Y., Isogai T., Sugano S.,

RT "NEDO human cDNA sequencing project."

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AK00502; BAA91209.1; -

SQ SEQUENCE 181 AA; 19669 MW; 4F82301D70322F5E CRC64;

## Alignment Scores:

Pred. No.: 2.83e-07 Length: 181  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.78% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x Q9NX13 (1-181)

QY 2035 GGTTCAGGATTCCTGCTGCTCAGCCCTCCCAAGTAGCTGGATTACA 1988

DB 150 GlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyIleThr 165

## RESULT 10

Q8TDM0 PRELIMINARY; PRT; 211 AA.

AC Q8TDM0; 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DE Breast carcinoma amplified sequence 4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Barlund M., Monni O., Weaver D.J., Kallioniemi A.;

RA "Cloning of two novel genes, BCAS3 and BCAS4, that are amplified,

RT overexpressed, and rearranged in breast cancer."

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF61220; AAL99633.1; -

SQ SEQUENCE 211 AA; 22758 MW; D16930DBC982AF5B CRC64;

## Alignment Scores:

Pred. No.: 2.76e-07 Length: 211  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.78% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x Q8TDM0 (1-211)

QY 2035 GGTTCAGGATTCCTGCTGCTCAGCCCTCCCAAGTAGCTGGATTACA 1988

DB 180 GlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyIleThr 195

## RESULT 11

Q9NR08 PRELIMINARY; PRT; 231 AA.

AC Q9NR08; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE Ubiquitous TPR-motif protein Y isoform (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=20319030; PubMed=10861003;

RA Shen P., Wang F., Underhill P.A., Franco C., Yang W.-H., Roxas A.,

RA Sung R., Lin A.A., Hyman R.W., Vollrath D., Davis R.W.,

RA Cavalli-Sforza L., Oefner P.J.;  
 RT "Population genetic implications from sequence variation in four Y  
 chromosome genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:7354-7359(2000).  
 DR EMBL: AF265575; AAF77052.1; -.  
 FT NON-TER  
 SQ SEQUENCE 231 AA; 25510 MW; 16111092D2244702 CRC64;

## Alignment Scores:

Pred. No.: 2.72e-07 Length: 221  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.78% Indels: 0  
 DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x Q9NR08 (1-231)

QY 1879 TCGGCTCCCAAGTGTGCGATTACAGCGTGAGCCAGCCCGG 1832  
 Db 216 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaArg 231

## RESULT 12

Q9H5R3 PRELIMINARY; PRT; 232 AA.

AC Q9H5R3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE CDNA: FLJ23147 fis, clone LINC09295.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
 RA Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y., Ota T., Suzuki Y.,  
 RA Ogasashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
 RA Isegaki T., Sugano S.;  
 RT "MEDO human cDNA sequencing project."  
 RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AK026800; BAB15557.1; -.  
 SQ SEQUENCE 232 AA; 26703 MW; 56946BE74C226FF5 CRC64;

## Alignment Scores:

Pred. No.: 2.72e-07 Length: 232  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.78% Indels: 0  
 DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x Q9H5R3 (1-232)

QY 2033 TTCACGCAATTCCTGCTGCTCAGCCTCCAGTAGCGGATTACAGG 1986  
 Db 186 PhelyArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArg 201

## RESULT 13

Q60448 PRELIMINARY; PRT; 375 AA.

ID Q60448;  
 AC Q60448;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Neuronal thread protein AD7c-NTP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=NEURONAL;  
 RA de la Monte S.M., Ghandari K., Frey W., Beheshti I., Hauser S.A.,  
 RA Ghandari H.A., Wands J.R.;  
 RT "Characterization of the AD7c-NTP cDNA and its expression in the brain  
 RT and cerebrospinal fluid of patients with Alzheimer's disease."  
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF010144; AAC08737.1; -.  
 SQ SEQUENCE 375 AA; 41720 MW; 955443950A5BFEDD CRC64;

## Alignment Scores:

Pred. No.: 2.52e-07 Length: 375  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.78% Indels: 0  
 DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x Q60448 (1-375)

QY 1879 TCGGCTCCCAAGTGTGCGATTACAGCGTGAGCCAGCCCGG 1832  
 Db 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaArg 294

## RESULT 14

Q9UHT2 PRELIMINARY; PRT; 126 AA.

AC Q9UHT2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE PRO1900.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Bl J., Zhang Y., Liu M.,  
 RA He F.;  
 RT "Functional prediction of the coding sequences of 33 new genes deduced  
 RT by analysis of cDNA clones from human fetal liver."  
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF118081; AAF22025.1; -.  
 SQ SEQUENCE 126 AA; 13439 MW; 183FEF16B1D25C91 CRC64;

## Alignment Scores:

Pred. No.: 3.51e-06 Length: 126  
 Score: 15.00 Matches: 15  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.67% Indels: 0  
 DB: 4 Gaps: 0

US-09-966-880a-7 (1-2818) x Q9UHT2 (1-126)

QY 2033 TCCTCGCTCAGCCTCCAGTAGCGGATTACAGGTCGCTCAGG 1979  
 Db 71 SerProAlaSerAlaSerGlnValAlaGlyIleThrGlyAlaCys 85

## RESULT 15

Q96LU6 PRELIMINARY; PRT; 166 AA.

ID Q96LU6;  
 AC Q96LU6;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CDNA FLJ25058 fis, clone CBL04608.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CEREBELLUM;  
 RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Suzuki Y.,  
 RA Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,  
 RA Sugiyama T., Irie K., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,  
 RA Kawakami B., Nagai K., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AK057787; BAB71572.1; -  
 SQ SEQUENCE 166 AA; 19376 MW; C9655E2C34C1FAF5 CRC64;

## Alignment Scores:

Pred. No.:	3.37e-06	Length:	166
Score:	15.00	Matches:	15
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.67%	Indels:	0
DB:	4	Gaps:	0

US-09-966-880A-7 (1-2818) x Q96LU6 (1-166)

QY 1998 GCTACTTGGAGGCTGAGGACAGAGATCGCTTGACACCGAGAGG 2042  
 Db 149 AlATnTTPGluAlaGluAlaGluGluSerLeuGluProArgArg 163

Search completed: July 7, 2003, 23:36:48  
 Job time : 111.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 7, 2003, 19:16:16 ; Search time 609 Seconds  
(without alignments)  
10420.577 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 agagacacatcataatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: OLIGO-MTC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 segs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2818	100.0	2818	21	AAC55312	Human activation-i
2	2174	77.1	6564	21	AAC55314	Human activation-i
3	2174	77.1	11204	21	AAC55339	Human activation-i
4	2172	77.1	2172	21	AAC55319	Human activation-i
5	1489	52.8	1665	22	AAK81088	Human immune/haema
6	574	20.4	574	22	AAK81089	Human immune/haema
7	429	15.2	429	22	AAK61819	Human immune/haema
8	271	9.6	271	21	AAC55317	Human activation-i
9	148	5.3	148	21	AAC55316	Human activation-i

10	87	3.1	87	21	AAC55315	Human activation-i
11	87	3.1	5514	21	AAC55313	Human activation-i
12	78	2.8	116	21	AAC55318	Human activation-i
13	74	2.6	280	22	AAK76441	Human immune/haema
14	70	2.5	481	23	ABV22611	Human prostate exp
15	70	2.5	481	23	ABV28433	Human prostate exp
16	70	2.5	529	23	ABV34298	Human prostate exp
17	70	2.5	529	23	ABV43160	Human prostate exp
18	70	2.5	533	22	AAK64557	Human immune/haema
19	70	2.5	564	22	ABV50819	Human prostate exp
20	70	2.5	577	23	ABV04010	Human prostate exp
21	70	2.5	781	23	ABV13179	Human prostate exp
22	70	2.5	2983	22	ABV15784	Human nervous syst
23	70	2.5	167343	24	ABL64403	Stomach cancer rel
24	70	2.5	167343	24	ABL67239	Human gutathione
25	67	2.4	50849	24	ABN87883	Human gutathione
26	66	2.3	125	22	AAK28488	Genomic sequence #
27	66	2.3	125	22	AAK28489	Genomic sequence #
28	66	2.3	451	22	ABL87490	Human ovarian canc
29	66	2.3	1674	22	AAK8482	Human polynucleoti
30	64	2.3	17216	22	ABA18237	Human nervous syst
31	64	2.3	17216	22	ABA17200	Human musculoskele
32	64	2.3	17217	22	ABA18239	Human nervous syst
33	64	2.3	17217	22	AAK37201	Human musculoskele
34	64	2.3	32249	22	ABL94931	Human reproductive
35	64	2.3	32249	22	ABL97825	Human testicular a
36	64	2.3	136284	24	ABK83575	Human CDNA differe
37	63	2.2	1995	22	AAH18425	Human CDNA sequenc
38	63	2.2	5350	24	ABK93137	Human prostate spe
39	63	2.2	8790	24	ABK69856	Human secreted pro
40	63	2.2	17730	22	AAK29826	Human cytoskeletal
41	63	2.2	27841	22	AAK29820	Human cytoskeletal
42	63	2.2	27841	22	AAK29827	Human cytoskeletal
43	63	2.2	31730	22	AAK37445	Human musculoskele
44	62	2.2	45000	22	AAK12437	DNA encoding 1-cam1
45	61	2.2	539	22	AAH08902	Human CDNA clone (

#### ALIGNMENTS

RESULT 1						
ID	AAC55312	standard; CDNA; 2818 BP.				
XX	XX					
AC	AAC55312;					
XX	XX					
DT	05-FEB-2001	(first entry)				
XX	XX					
DE	Human activation-induced cytidine deaminase encoding CDNA SEQ ID NO:7.					
XX	XX					
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;					
KW	immune related disease; allergy; allergic disease; antileptotic;					
KW	antileptotic; antileptotic; ophthalmological; anti-HIV; dermatological;					
KW	gene therapy; B cell associated immune system disorder; food allergy;					
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;					
KW	IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;					
KW	drug allergy; allergic rhinitis; Rosen disease; Digorge disease; ARDS;					
KW	ataxia telangiectasia; common variable immunodeficiency disorder;					
KW	major histocompatibility class II deficiency disease;					
KW	auto immunodeficiency syndrome; Igg subclass selection disorder; ss.					
OS	Homo sapiens.					
XX	XX					
EH	Key	Location/Qualifiers				
FT	CDS	80..676				
FT	FT	/tag= a				
FT	FT	/product= "activation-induced cytidine deaminase"				
XX	XX					
PN	WO200058480-A1.					
XX	XX					
PD	05-OCT-2000.					
XX	XX					



PF 28-MAR-2000; 2000MO-JF01918.  
XX  
XX 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
PA (NIBS) JAPAN TOBACCO INC.  
PA (HONJ) HONJO T.  
PI Honjo T, Muramatsu M.  
XX  
XX WPI; 2000-611715/58.  
DR P-PSDB; AAB24198.  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
XX Claim 3; Page 135-139; 174pp; Japanese.  
XX  
CC The present sequence encodes human activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,  
CC antitumoric, antiasthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, discoid disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IGE disorder, and Igg subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders.  
XX  
SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;

Query Match 100.0%; Score 2818; DB 21; Length 2818;  
Best Local Similarity 100.0%; Pred No. 0;  
Matches 2818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGAACCATCATTAATTAAGTATGATTTTCTGGCCCTGAGACTTGACGAGGAGCAGA 60  
DB 1 AGAGAACCATCATTAATTAAGTATGATTTTCTGGCCCTGAGACTTGACGAGGAGCAGA 60  
QY 61 AGACACTGTGACACACCATATGACACGCTCTGTATGAACCGAGAAAGTTCTTTACCA 120  
DB 61 AGACACTGTGACACACCATATGACACGCTCTGTATGAACCGAGAAAGTTCTTTACCA 120  
QY 121 ATTCAAAATATGTCGCTGGGCTAAGGCTGAGGCTGAGACCTACCTGTGTACGTAGTAA 180  
DB 121 ATTCAAAATATGTCGCTGGGCTAAGGCTGAGGCTGAGACCTACCTGTGTACGTAGTAA 180  
QY 181 GAGGCGTGAAGTGTACATCTTTCACTGACCTTGTATCTTCGGAATAAAGAGG 240  
DB 181 GAGGCGTGAAGTGTACATCTTTCACTGACCTTGTATCTTCGGAATAAAGAGG 240  
QY 241 CTGCGACGTGAATGCTCTCTCCGCTACATCTCGAGCTGGAGACTGAGACCTGAGCG 300  
DB 241 CTGCGACGTGAATGCTCTCTCCGCTACATCTCGAGCTGGAGACTGAGACCTGAGCG 300  
QY 301 CTGCTACCGCGTACCTGCTGCTGCTGCTGAGACCCCTGCTACGACTGTGCCGACATGT 360  
DB 301 CTGCTACCGCGTACCTGCTGCTGCTGCTGAGACCCCTGCTACGACTGTGCCGACATGT 360  
QY 361 GGGCGACTTTTGGCGAGGAGAACCCCAACTCACTGAGAGATCTTACCGCGGCTCTTA 420  
DB 361 GGGCGACTTTTGGCGAGGAGAACCCCAACTCACTGAGAGATCTTACCGCGGCTCTTA 420  
QY 421 CTTCCTGAGAGACCGGAAGGTGAGCCGAGGAGGCTGGGCGGTGACACCGCGCGGGGT 480  
DB 421 CTTCCTGAGAGACCGGAAGGTGAGCCGAGGAGGCTGGGCGGTGACACCGCGCGGGGT 480

DB 421 CTTCCTGAGAGACCGGAAGGTGAGCCGAGGAGGCTGGGCGGTGACACCGCGCGGGGT 480  
QY 481 GCAATAGCCATCAATGACCTTCAAGATATATTTTACTGCTGGAATCTTTTGAAGAAA 540  
DB 481 GCAATAGCCATCAATGACCTTCAAGATATATTTTACTGCTGGAATCTTTTGAAGAAA 540  
QY 541 CCATGAAGAAGACTTCAAGAGCTGGAGAGGCTGCATGAATAATCACTGCTGCTCAG 600  
DB 541 CCATGAAGAAGACTTCAAGAGCTGGAGAGGCTGCATGAATAATCACTGCTGCTCAG 600  
QY 601 AAGCTTGGCGCATCCTTTTGCCCTGTATGAGGTGATGACTTACGAGACCATTTTCG 660  
DB 601 AAGCTTGGCGCATCCTTTTGCCCTGTATGAGGTGATGACTTACGAGACCATTTTCG 660  
QY 661 TACCTTGGAGATTGATGACCACTTCCAGAAATCTCACACAGTGAATAATCTGCTG 720  
DB 661 TACCTTGGAGATTGATGACCACTTCCAGAAATCTCACACAGTGAATAATCTGCTG 720  
QY 721 AAGACAGTGAATAAAAACAGTCTCAAGCTCTCTGTTTATCTTCAACTCTCAC 780  
DB 721 AAGACAGTGAATAAAAACAGTCTCTCAAGCTCTCTGTTTATCTTCAACTCTCAC 780  
QY 781 TTTCTTACAGTTTACAGAAAAAATTTATATACGACTTTTAAAAAGATCTATGCTTG 840  
DB 781 TTTCTTACAGTTTACAGAAAAAATTTATATACGACTTTTAAAAAGATCTATGCTTG 840  
QY 841 AAATAGAGAGAACACAGCTCTGGCCAGGAGAGCTCTGCAATGGTGCAGTTTGAAT 900  
DB 841 AAATAGAGAGAACACAGCTCTGGCCAGGAGAGCTCTGCAATGGTGCAGTTTGAAT 900  
QY 901 GCAACATGTCTCCCTACAGGAATACAGAACTCAGACCTGGAGACCTCTAAAGTGT 960  
DB 901 GCAACATGTCTCCCTACAGGAATACAGAACTCAGACCTGGAGACCTCTAAAGTGT 960  
QY 961 CAACGTTTCTATGACTTTAGGTAGAGAGAGAGAGTATCTCTAAAAAGCATG 1020  
DB 961 CAACGTTTCTATGACTTTAGGTAGAGAGAGAGTATCTCTAAAAAGCATG 1020  
QY 1021 GTGAGAGATTAATGTTTATATATCAACATCTTTATTTATGATTCATTTAGTTTAC 1080  
DB 1021 GTGAGAGATTAATGTTTATATATCAACATCTTTATTTATGATTCATTTAGTTTAC 1080  
QY 1081 AGTGGTGTAGTATGATTTTCTATCTTTCTCTGAGCTTACTTCAAGTAAACAC 1140  
DB 1081 AGTGGTGTAGTATGATTTTCTATCTTTCTCTGAGCTTACTTCAAGTAAACAC 1140  
QY 1141 AAACCTTCCATCAGGACATGATCTATAGACCTCTCTATAGAGATCTCTGGGATGT 1200  
DB 1141 AAACCTTCCATCAGGACATGATCTATAGACCTCTCTATAGAGATCTCTGGGATGT 1200  
QY 1201 GACCCCAACCATCTCTCCAAAGCATTAATATCAATCAAGCCGTGATATTTTAAACAG 1260  
DB 1201 GACCCCAACCATCTCTCCAAAGCATTAATATCAATCAAGCCGTGATATTTTAAACAG 1260  
QY 1261 CAGAGACATGTTTATGTTTGTACAAAGAAATTTTATGGGTGGGATGAGAGTATA 1320  
DB 1261 CAGAGACATGTTTATGTTTGTACAAAGAAATTTTATGGGTGGGATGAGAGTATA 1320  
QY 1321 GACCATGATGATGATCTTCAAGTACTTAAATTAAGATCTTAAATGGGAGAGAGAC 1380  
DB 1321 GACCATGATGATGATCTTCAAGTACTTAAATTAAGATCTTAAATGGGAGAGAGAC 1380  
QY 1381 TGTGAACAGACACCTTAATATATGATGATGATGATGATGATGATGATGATGATGAT 1440  
DB 1381 TGTGAACAGACACCTTAATATATGATGATGATGATGATGATGATGATGATGATGAT 1440  
QY 1441 AAACCTTTTAAAGAAAGTCCCTATTTAGAAACACCAACCACTTACATATATATTA 1500  
DB 1441 AAACCTTTTAAAGAAAGTCCCTATTTAGAAACACCAACCACTTACATATATATTA 1500  
QY 1501 GCAACAAATGGAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
DB 1501 GCAACAAATGGAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1560

QY 1561 GGGCTCTTATCTCAGAAAAAGCCATCAGTCAAGTTGCTACATTTTGTATGTGTG 1620  
Db 1561 GGGCTCTTATCTCAGAAAAAGCCATCAGTCAAGTTGCTACATTTTGTATGTGTG 1620  
QY 1621 GATGCTCTCTCCCAAGGATATATTAAGATATTAAGAGAGTTGTACAAAACAGATATAA 1680  
Db 1621 GATGCTCTCTCCCAAGGATATATTAAGATATTAAGAGAGTTGTACAAAACAGATATAA 1680  
QY 1681 AGCTGCGAACCCTGGGACAGCTCATCTAGCTGTGGAGAGTTGAGAGAGGAGGA 1740  
Db 1681 AGCTGCGAACCCTGGGACAGCTCATCTAGCTGTGGAGAGTTGAGAGAGGAGGA 1740  
QY 1741 TGGCTTGAACACAGTGTCTTCAAGCCAGCTGGGCAACATTAACAAGATCTGTCTCAA 1800  
Db 1741 TGGCTTGAACACAGTGTCTTCAAGCCAGCTGGGCAACATTAACAAGATCTGTCTCAA 1800  
QY 1801 AAAAAAAAAAAAAAAAAAGAGAGAGGCGGCGTGTGTGCTCAGCGCTGTATCC 1860  
Db 1801 AAAAAAAAAAAAAAAAAAGAGAGAGGCGGCGTGTGTGCTCAGCGCTGTATCC 1860  
QY 1861 CAGACTTTGGAGGCGGCGGCGGATCAGCTGTGTGCTCAGAGATTTGAGACACCT 1920  
Db 1861 CAGACTTTGGAGGCGGCGGCGGATCAGCTGTGTGCTCAGAGATTTGAGACACCT 1920  
QY 1921 GGCCACATGCAAAACCCGCTGTACTCAAAATGCAAAATTTAGCCAGGCGTGTAGC 1980  
Db 1921 GGCCACATGCAAAACCCGCTGTACTCAAAATGCAAAATTTAGCCAGGCGTGTAGC 1980  
QY 1981 AGGCACTGTATCCAGCTACTTGGAGGCGTGGAGAGAGATGCTTGAACCCAGGA 2040  
Db 1981 AGGCACTGTATCCAGCTACTTGGAGGCGTGGAGAGAGATGCTTGAACCCAGGA 2040  
QY 2041 GGTGGAGGTTGCAAGTGAAGTGTGCGCTGTCACTCCAGCTGGGCGACAGAGCA 2100  
Db 2041 GGTGGAGGTTGCAAGTGAAGTGTGCGCTGTCACTCCAGCTGGGCGACAGAGCA 2100  
QY 2101 AGACTCTGTCTCAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 2160  
Db 2101 AGACTCTGTCTCAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 2160  
QY 2161 GGAAGAGAGAGTGGGAGAGATTCAGAGAAATGCTTTATCCAAACAAATGTAAAGA 2220  
Db 2161 GGAAGAGAGAGTGGGAGAGATTCAGAGAAATGCTTTATCCAAACAAATGTAAAGA 2220  
QY 2221 GCCAATTAAGGATCCCTATTGCTCTTTTGTGTCTATTGTCCTTAAACATGCTCTT 2280  
Db 2221 GCCAATTAAGGATCCCTATTGCTCTTTTGTGTCTATTGTCCTTAAACATGCTCTT 2280  
QY 2281 GACAGTGAAGAAAAATTCAGATTAACCATATCCCTGTCCTTATTAAGCAACCT 2340  
Db 2281 GACAGTGAAGAAAAATTCAGATTAACCATATCCCTGTCCTTATTAAGCAACCT 2340  
QY 2341 TGCATTAAGAGATGAGAGATCCACAGAAAACTTGAATGCACACAGCTCTTTTAAATC 2400  
Db 2341 TGCATTAAGAGATGAGAGATCCACAGAAAACTTGAATGCACACAGCTCTTTTAAATC 2400  
QY 2401 TTAATGTACATAAGTTGTAAGAGTAAAAATTTACTATCATGTATTCATTTATAT 2460  
Db 2401 TTAATGTACATAAGTTGTAAGAGTAAAAATTTACTATCATGTATTCATTTATAT 2460  
QY 2461 TTAATGTATTTTCCGCTATATGATTTTATTAACATGATTCCTTTTGTATTAATGA 2520  
Db 2461 TTAATGTATTTTCCGCTATATGATTTTATTAACATGATTCCTTTTGTATTAATGA 2520  
QY 2521 AATGAGATCTCAAGGCTTCATTAATTTATTAATTTTGAATGATTTTAAACAGCT 2580  
Db 2521 AATGAGATCTCAAGGCTTCATTAATTTATTAATTTTGAATGATTTTAAACAGCT 2580  
QY 2581 GTAATTTGAACATTTGAGATTAAGTGTCTGAGAGCAATTTCTGTGATTTTAACT 2640  
Db 2581 GTAATTTGAACATTTGAGATTAAGTGTCTGAGAGCAATTTCTGTGATTTTAACT 2640

QY 2641 TTTATGACAGCAAAATTTGCTTCTGCTCAGCTTCAATCAGTTAAATGATAAATAT 2700  
Db 2641 TTTATGACAGCAAAATTTGCTTCTGCTCAGCTTCAATCAGTTAAATGATAAATAT 2700  
QY 2701 TTTGAGAGCTGTGAAGATTAATTAACCAATTAATTAATTAAGTATTAATGAAGT 2760  
Db 2701 TTTGAGAGCTGTGAAGATTAATTAACCAATTAATTAATTAAGTATTAATGAAGT 2760  
QY 2761 TAAATTAATAATTCAGATGATGATTAATTAATTAATTAATTAATTAATTAATTA 2818  
Db 2761 TAAATTAATAATTCAGATGATGATTAATTAATTAATTAATTAATTAATTAATTA 2818  
RESULT 2  
AAC55314  
ID AAC55314 standard; DNA; 6564 BP.  
AAC55314;  
05-FEB-2001 (first entry)  
XX  
XX  
DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.  
XX  
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianemic; antistimatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; digeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; Igg subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200058480-A1.  
PD  
XX  
XX 05-OCT-2000.  
PF  
XX 28-MAR-2000; 2000WO-JP01918.  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
PA (NIBS) JAPAN TOBACCO INC.  
PA (HONJ) HONJO T.  
XX  
XX Honjo T, Muramatsu M;  
PI WPI; 2000-611715/58.  
DR  
XX  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
PS  
XX  
XX  
PS Claim 17; Page 145-150; 174pp; Japanese.  
XX  
XX  
CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianemic, antistimatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (Iga) deficiency  
CC disease, Iga nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, digeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC Ige disorder, and Igg subclass selection disorder. The DNA sequences







CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents the exon 5 genomic DNA sequence of human AID.  
XX  
SQ Sequence 2172 BP; 702 A; 379 C; 465 G; 626 T; 0 other;

Query Match 77.18; Score 2172; DB 21; Length 2172;  
Best Local Similarity 100.0%; Pred. No. 0;

Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 CCCCCTATGAGTGTATGACTTACGACGATTTCCGACTTTGGGACTTTGATGCAA 682  
DB 1 CCCCCTATGAGTGTATGACTTACGACGATTTCCGACTTTGGGACTTTGATGCAA 60  
QY 683 CTTCAGGAATGTCACACAGATGAATATCTCTGCTGAAGACAGAGGATAAACAACT 742  
DB 61 CTTCAGGAATGTCACACAGATGAATATCTCTGCTGAAGACAGAGGATAAACAACT 120  
QY 743 CCTTCAAGTCTCTGTTTTTATTTCTTCACTCTCACTTTCTTAGAGTTTACAGAAAA 802  
DB 121 CCTTCAAGTCTCTGTTTTTATTTCTTCACTCTCACTTTCTTAGAGTTTACAGAAAA 180  
QY 803 AATTATATACACACTCTTAAAAATCTATGCTTGAATAATAGAGAGAACACAGCT 862  
DB 181 AATTATATACACACTCTTAAAAATCTATGCTTGAATAATAGAGAGAACACAGCT 240  
QY 863 CTGGCCAGGAGCTGCTGCAATGTCAGATTTTGCAATGCAACATTTGCCCTACGGGA 922  
DB 241 CTGGCCAGGAGCTGCTGCAATGTCAGATTTTGCAATGCAACATTTGCCCTACGGGA 300  
QY 923 ATAAACAAGCTGAGAGACCTGGAGACCTCTAAAGTGTCAAGTTTTCTATGACTTTTA 982  
DB 301 ATAAACAAGCTGAGAGACCTGGAGACCTCTAAAGTGTCAAGTTTTCTATGACTTTTA 360  
QY 983 GGTAGATGAGACCAAGGTAGATCTTAAAAAGCTGTGTGAGAGATTAATGTTTTTA 1042  
DB 361 GGTAGATGAGACCAAGGTAGATCTTAAAAAGCTGTGTGAGAGATTAATGTTTTTA 420  
QY 1043 TATCAACATCCTTATTTATTTGATTCATTTGATTTAGAGTGTGTAGTATGATTT 1102  
DB 421 TATCAACATCCTTATTTATTTTATTTGATTTGATTTAGAGTGTGTAGTATGATTT 480  
QY 1103 TCTATCTTTTCCCTTGAAGTTACTTTCAGTAACACAACTCTTCCATCGAGCCATA 1162  
DB 481 TCTATCTTTTCCCTTGAAGTTACTTTCAGTAACACAACTCTTCCATCGAGCCATA 540  
QY 1163 TCTATGAGACCTCTTATGAGATCTGAGTGTGATTTGTGACCCCAACATCTCTCCAA 1222  
DB 541 TCTATGAGACCTCTTATGAGATCTGAGTGTGATTTGTGACCCCAACATCTCTCCAA 600  
QY 1223 GCATTATATCCCAATATGCGCTGATGTTTATATAGACAGAAAGCTGTTTTATGTTG 1282  
DB 601 GCATTATATCCCAATATGCGCTGATGTTTATATAGACAGAAAGCTGTTTTATGTTG 660  
QY 1283 TACAAAAAGAAAGATTGTATGGGTGGAGTGAAGGTATAGCCATGATGATGATCTCAA 1342  
DB 661 TACAAAAAGAAAGATTGTATGGGTGGAGTGAAGGTATAGCCATGATGATGATCTCAA 720  
QY 1343 GCTACTTTTATAAGGATCTTAAATGGGAGAGAGACTGTAAACAGACACCTATATA 1402  
DB 721 GCTACTTTTATAAGGATCTTAAATGGGAGAGAGACTGTAAACAGACACCTATATA 780  
QY 1403 TGGGTGATGCTGGAAGTACCAATCTTGTGAAACGCAAACTCTTTTAAAGAGTCCCT 1462  
DB 781 TGGGTGATGCTGGAAGTACCAATCTTGTGAAACGCAAACTCTTTTAAAGAGTCCCT 840  
QY 1463 AATTGAAACACCCCAAACTTCACATATCATATTAATAGCAACAAATTTGAAGGAAGTTG 1522  
DB 841 AATTGAAACACCCCAAACTTCACATATCATATTAATAGCAACAAATTTGAAGGAAGTTG 900  
QY 1523 CTGGAATGTTGGGAGAGAAATCTATTGGCTCTGCTGGGTCTCTTCATCTCAGAAATG 1582  
DB 901 CTGGAATGTTGGGAGAGAAATCTATTGGCTCTGCTGGGTCTCTTCATCTCAGAAATG 960

QY 1583 CCAATCAGTCAAGTTGCTACATTTGTATGTGTGATGCTTCTCCAAAGATATAT 1642  
DB 961 CCAATCAGTCAAGTTGCTACATTTGTATGTGTGATGCTTCTCCAAAGATATAT 1020  
QY 1643 TAACTATATAGAGATTTGACAAAACAGAAATGATAAAGCTGGCAACCGTGACACGC 1702  
DB 1021 TAACTATATAGAGATTTGACAAAACAGAAATGATAAAGCTGGCAACCGTGACACGC 1080  
QY 1703 TCATAGTCTAGCTGCTTGGGAGGTGAGAGGAGAGATGCTTGAACACAGGTTCATA 1762  
DB 1081 TCATAGTCTAGCTGCTTGGGAGGTGAGAGGAGAGATGCTTGAACACAGGTTCATA 1140  
QY 1763 GAGCCAGCTGGGCAACATTAACAAGTCTGCTCTCAAAAAAAGAAAAAAGAAA 1822  
DB 1141 GAGCCAGCTGGGCAACATTAACAAGTCTGCTCTCAAAAAAAGAAAAAAGAAA 1200  
QY 1823 GAGAGAGGCGGGGCTGTGCTCACGCCGTATATCCACACTTTGGAGGCCGAGCC 1882  
DB 1201 GAGAGAGGCGGGGCTGTGCTCACGCCGTATATCCACACTTTGGAGGCCGAGCC 1260  
QY 1883 GGGCGGATCACCTGTGTGTCAGGAGTTTGAGACCAAGCTGGCCAAACATGGCAAAACCCGT 1942  
DB 1261 GGGCGGATCACCTGTGTGTCAGGAGTTTGAGACCAAGCTGGCCAAACATGGCAAAACCCGT 1320  
QY 1943 CTGTACTCAAAATGCAAAATTTAGCCAGGGCTGTAGAGGCACTGTAAATCCAGCTAC 2002  
DB 1321 CTGTACTCAAAATGCAAAATTTAGCCAGGGCTGTAGAGGCACTGTAAATCCAGCTAC 1380  
QY 2003 TTGGAGAGCTGAGGAGAGATGCTTGAACCCAGAGAGGTGAGAGTTGCAAGTGA 2062  
DB 1381 TTGGAGAGCTGAGGAGAGATGCTTGAACCCAGAGAGGTGAGAGTTGCAAGTGA 1440  
QY 2063 GATCGTCCGTTGCACTCCAGCTGGGCGACAGACAAAGCTGTGCTCAGAAAAAAA 2122  
DB 1441 GATCGTCCGTTGCACTCCAGCTGGGCGACAGACAAAGCTGTGCTCAGAAAAAAA 1500  
QY 2123 AAAAAAAGAGAGAGAGAGAGAGAGAGAAATTTGGGAGAGAGAGAGAGAGAGAT 2182  
DB 1501 AAAAAAAGAGAGAGAGAGAGAGAGAGAAATTTGGGAGAGAGAGAGAGAGAGAT 1560  
QY 2183 TGCAGAGAAATTTGCTTATCCAAACAAATGTAGAGAGCCAAATAGGATCCCTATTTG 2242  
DB 1561 TGCAGAGAAATTTGCTTATCCAAACAAATGTAGAGAGCCAAATAGGATCCCTATTTG 1620  
QY 2243 TCTCTTTTGTGTCTATTTGTCCCTAACACTGTCTTGAAGTGAAGAAAAATTTGAGA 2302  
DB 1621 TCTCTTTTGTGTCTATTTGTCCCTAACACTGTCTTGAAGTGAAGAAAAATTTGAGA 1680  
QY 2303 ATAAACAATGCCCTGTGCTTATTTACCTAGCAACCTTGCATGAATGAAGATGAGATCC 2362  
DB 1681 ATAAACAATGCCCTGTGCTTATTTACCTAGCAACCTTGCATGAAGATGAGATCC 1740  
QY 2363 ACAGAAAACTGATGACACAAGTGTATTTATATCTTTTACATTAAGTTGATA 2422  
DB 1741 ACAGAAAACTGATGACACAAGTGTATTTATTAATCTTTTACATTAAGTTGATA 1800  
QY 2423 AGAGTTAAAAATTTGTTACTCATGTATTTATTTATTTATTTATTTTTCGCTATG 2482  
DB 1801 AGAGTTAAAAATTTGTTACTCATGTATTTATTTATTTATTTATTTTTCGCTATG 1860  
QY 2483 ATTTTTATTAACATGATTTCCCTTCTGATATATGAAATGAGTCCAAAGCTCATA 2542  
DB 1861 ATTTTTATTAACATGATTTCCCTTCTGATATATGAAATGAGTCCAAAGCTCATA 1920  
QY 2543 AATTATATCTTTGAAATGATTTCTATTAACAAGATGTAAATGTAAATGTGAGTAAT 2602  
DB 1921 AATTATATCTTTGAAATGATTTCTATTAACAAGATGTAAATGTAAATGTGAGTAAT 1980  
QY 2603 GGTGTACGAGACATTTCTGTATTTTGTAACTTTATAGACGAAATTTGCTTC 2662  
DB 1981 GGTGTACGAGACATTTCTGTATTTTGTAACTTTATAGACGAAATTTGCTTC 2040

PR	01-SEP-2000	2000US-0229345	
PR	05-SEP-2000	2000US-0229509	
PR	05-SEP-2000	2000US-0229513	
PR	06-SEP-2000	2000US-0230437	
PR	06-SEP-2000	2000US-0230438	
PR	06-SEP-2000	2000US-0231242	
PR	08-SEP-2000	2000US-0231243	
PR	08-SEP-2000	2000US-0231244	
PR	08-SEP-2000	2000US-0231414	
PR	08-SEP-2000	2000US-0231415	
PR	08-SEP-2000	2000US-0232080	
PR	12-SEP-2000	2000US-0232081	
PR	12-SEP-2000	2000US-0231968	
PR	14-SEP-2000	2000US-0232397	
PR	14-SEP-2000	2000US-0232398	
PR	14-SEP-2000	2000US-0232399	
PR	14-SEP-2000	2000US-0232400	
PR	14-SEP-2000	2000US-0232401	
PR	14-SEP-2000	2000US-0233063	
PR	14-SEP-2000	2000US-0233064	
PR	14-SEP-2000	2000US-0233065	
PR	21-SEP-2000	2000US-0234223	
PR	21-SEP-2000	2000US-0234224	
PR	25-SEP-2000	2000US-0234997	
PR	25-SEP-2000	2000US-0234998	
PR	26-SEP-2000	2000US-0235484	
PR	27-SEP-2000	2000US-0235484	
PR	27-SEP-2000	2000US-0235896	
PR	29-SEP-2000	2000US-0236337	
PR	29-SEP-2000	2000US-0236337	
PR	29-SEP-2000	2000US-0236368	
PR	29-SEP-2000	2000US-0236369	
PR	29-SEP-2000	2000US-0236370	
PR	02-OCT-2000	2000US-0236802	
PR	02-OCT-2000	2000US-0237037	
PR	02-OCT-2000	2000US-0237038	
PR	02-OCT-2000	2000US-0237039	
PR	02-OCT-2000	2000US-0237040	
PR	13-OCT-2000	2000US-0239935	
PR	13-OCT-2000	2000US-0239937	
PR	20-OCT-2000	2000US-0240960	
PR	20-OCT-2000	2000US-0241221	
PR	20-OCT-2000	2000US-0241785	
PR	20-OCT-2000	2000US-0241786	
PR	20-OCT-2000	2000US-0241787	
PR	20-OCT-2000	2000US-0241808	
PR	20-OCT-2000	2000US-0241816	
PR	01-NOV-2000	2000US-0244617	
PR	08-NOV-2000	2000US-0246674	
PR	08-NOV-2000	2000US-0246675	
PR	08-NOV-2000	2000US-0246676	
PR	08-NOV-2000	2000US-0246677	
PR	08-NOV-2000	2000US-0246678	
PR	08-NOV-2000	2000US-0246528	
PR	08-NOV-2000	2000US-0246532	
PR	08-NOV-2000	2000US-0246539	
PR	08-NOV-2000	2000US-0246540	
PR	08-NOV-2000	2000US-0246541	
PR	08-NOV-2000	2000US-0246542	
PR	08-NOV-2000	2000US-0246543	
PR	08-NOV-2000	2000US-0246544	
PR	08-NOV-2000	2000US-0246545	
PR	08-NOV-2000	2000US-0246546	
PR	08-NOV-2000	2000US-0246547	
PR	08-NOV-2000	2000US-0246548	
PR	08-NOV-2000	2000US-0246549	
PR	08-NOV-2000	2000US-0246550	
PR	08-NOV-2000	2000US-0246551	
PR	08-NOV-2000	2000US-0246552	
PR	08-NOV-2000	2000US-0246553	
PR	08-NOV-2000	2000US-0246554	
PR	08-NOV-2000	2000US-0246555	
PR	08-NOV-2000	2000US-0246556	
PR	08-NOV-2000	2000US-0246557	
PR	08-NOV-2000	2000US-0246558	
PR	08-NOV-2000	2000US-0246559	
PR	08-NOV-2000	2000US-0246560	
PR	08-NOV-2000	2000US-0246561	
PR	08-NOV-2000	2000US-0246562	
PR	08-NOV-2000	2000US-0246563	
PR	08-NOV-2000	2000US-0246564	
PR	08-NOV-2000	2000US-0246565	
PR	08-NOV-2000	2000US-0246566	
PR	08-NOV-2000	2000US-0246567	
PR	08-NOV-2000	2000US-0246568	
PR	08-NOV-2000	2000US-0246569	
PR	08-NOV-2000	2000US-0246570	
PR	08-NOV-2000	2000US-0246571	
PR			







DB 1437 CGGGCGGATCACCCTGGTGGTCAGAGATTAGACCACTGGCCAACTGGCAAAACCCCG 1496  
QY 1942 TCTGTACTCAAAATGCAAAATTTAGCCAGCGCTGTAGCAGCACCTGTATATCCAGCTA 2001  
DB 1497 TCTGTACTCAAAATGCAAAATTTAGCCAGCGCTGTAGCAGCACCTGTATATCCAGCTA 1556  
QY 2002 CTTGGGAGGCTGAGGAGAGAGATGCTTGAACCCAGAGAGTTGACAGTGAAGCTG 2061  
DB 1557 CTTGGGAGGCTGAGGAGAGAGATGCTTGAACCCAGAGAGTTGACAGTGAAGCTG 1616  
QY 2062 AGATCGTGGCGTTGACCTCCAGCTGGGAGACAGAGCAAGCTCTGTC 2110  
DB 1617 AGATCGTGGCGTTGACCTCCAGCTGGGAGACAGAGCAAGCTCTGTC 1665  
RESULT 6  
AAK81089  
ID AAK81089 standard; DNA; 574 BP.  
AC AAK81089;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35901.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytosolic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
PN W0200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUL-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226861.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230457.  
PR 06-SEP-2000; 2000US-0230458.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236337.  
PR 29-SEP-2000; 2000US-0236337.  
PR 29-SEP-2000; 2000US-0236338.  
PR 29-SEP-2000; 2000US-0236339.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239355.  
PR 13-OCT-2000; 2000US-0239357.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246535.  
PR 08-NOV-2000; 2000US-0246536.  
PR 08-NOV-2000; 2000US-0246537.  
PR 08-NOV-2000; 2000US-0246537.  
PR 08-NOV-2000; 2000US-0246538.  
PR 08-NOV-2000; 2000US-0246538.  
PR 08-NOV-2000; 2000US-0246539.  
PR 08-NOV-2000; 2000US-0246539.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251088.  
PR 05-DEC-2000; 2000US-0251869.  
PR 06-DEC-2000; 2000US-0251879.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251899.  
PR 08-DEC-2000; 2000US-0251900.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-483426/52.  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX  
PS Disclosure; SEQ ID NO 35901; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK5950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 574 BP; 201 A; 80 C; 85 G; 208 T; 0 other;  
SQ

Query Match 20.4%; Score 574; DB 22; Length 574;  
Best Local Similarity 100.0%; Pred No. 2.8e-202; Indels 0; Gaps 0;  
Matches 574; Conservative 0; Mismatches 0;

QY 2211 AATGTAAGAGCAATTAAGGATCCCTATTGCTCTTTGCTGCTATTGCTCCCTAAC 2270  
DB 1 AATGTAAGAGCAATTAAGGATCCCTATTGCTCTTTGCTGCTATTGCTCCCTAAC 60  
QY 2271 AACTGCTTTGACAGTGAGAAAATTTTCAGAAATACCATTCCTGTCCTATTATAC 2330  
DB 61 AACTGCTTTGACAGTGAGAAAATTTTCAGAAATACCATTCCTGTCCTATTATAC 120  
QY 2331 TACCAACCCCTTGAAATGAAGATGAGCAATCCACAGGAAGAACTTGATGACAACTGCT 2390

DB 121 TAGCAACCCCTTGCAATGAAGATGAGCAATCCACAGGAAGAACTTGAAATGACAACTGCT 180  
QY 2391 TATTTAATCTATTGTACATAGTTGTAAAGATTAATAATGTTACTCATGTAT 2450  
DB 181 TATTTAATCTATTGTACATAGTTGTAAAGATTAATAATGTTACTCATGTAT 240  
QY 2451 CATTTATATTTATATTTATTTTGGCTCAATGATTTTATTAACATGATTTCTTTCT 2510  
DB 241 CATTTATATTTATATTTATTTTGGCTCAATGATTTTATTAACATGATTTCTTTCT 300  
QY 2511 GATATATTTGAAATGAGAGCTCAAAAGCTTCATTAATTTAATCTTAGAATGATTCAT 2570  
DB 301 GATATATTTGAAATGAGAGCTCAAAAGCTTCATTAATTTAATCTTAGAATGATTCAT 360  
QY 2571 AACCAAGTATGTAATGTGAATGCAATATGCTGCTACGACACCATTTCTGTGATT 2630  
DB 361 AACCAAGTATGTAATGTGAATGCAATATGCTGCTACGACACCATTTCTGTGATT 420  
QY 2631 TTAGTAACTTTTATGACAGCAAAATTTGCTTGGCTCAGCTTCATTCAGTAAATTAAT 2690  
DB 421 TTAGTAACTTTTATGACAGCAAAATTTGCTTGGCTCAGCTTCATTCAGTAAATTAAT 480  
QY 2691 GATAAATATTTTGGAGAGCTGTGAAGATTAATACCAATTAATTAATTAATTAATGATT 2750  
DB 481 GATAAATATTTTGGAGAGCTGTGAAGATTAATACCAATTAATTAATTAATTAATGATT 540  
QY 2751 TATATGAAGTTAAATTAATAAATTCAGTATGATG 2784  
DB 541 TATATGAAGTTAAATTAATAAATTCAGTATGATG 574

RESULT 7  
AAK61819  
ID AAK61819 standard; CDNA: 429 BP.  
XX  
AC AAK61819;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
XX Human immune/hematopoietic antigen encoding cDNA SEQ ID NO:6879.  
DE  
XX  
XX Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216880.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.





KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Disgeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
OS Homo sapiens.  
PN WO200058480-A1.  
PD 05-OCT-2000.  
PE 28-MAR-2000; 2000WO-JP01918.  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX (NISR ) JAPAN TOBACCO INC.  
PA (HONJ/) HONJO T.  
PI Honjo T, Muramatsu M;  
PI WPI; 2000-611715/58.  
DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX Claim 18; Page 150; 174pp; Japanese.  
XX The present invention describes an activation-induced cytidine deaminase  
XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
XX has cytidine activity similar to APOBEC-1. AID has anti-allergic,  
XX antianemic, antiasthmatic, ophthalmological, anti-HIV and  
XX dermatological activities, and can be used in gene therapy. AID  
XX polynucleotides are useful in methods for identifying drugs for the  
XX treatment of B cell associated immune system disorders, immunodeficiency  
XX diseases and allergies, such as immunoglobulin A (IGA) deficiency  
XX disease, IGA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
XX colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
XX disease, Disgeorge disease, ataxia telangiectasia, common variable  
XX immunodeficiency disorder, MHC (major histocompatibility class) class  
XX II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
XX IgG disorder, and IgG subclass selection disorder. The DNA sequences  
XX encoding AID may be used for gene therapy and the antibodies to the AID  
XX protein may be used for diagnosis and treatment of these disorders. The  
XX present sequence represents the exon 2 genomic DNA sequence of human AID.  
SQ Sequence 148 BP; 35 A; 33 C; 38 G; 42 T; 0 other;  
Query Match 5.3%; Score 148; DB 21; Length 148;  
Best Local Similarity 100.0%; Pred. No. 1.5e+46;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 CCTCTTGATGAACCGAGAGAGTTTCTTACCAATTCATAAATGTCGGCTAGGTAAGG 147  
1.CCTCTTGATGAACCGAGAGAGTTTCTTACCAATTCATAAATGTCGGCTAGGTAAGG 60  
DB 1.TCGGCGTGAAGACTACCTGCTGCTAGCTAGTGAAGAGCGCTGACAGTCTACTCCTTTC 207  
QY 148 TCGGCGTGAAGACTACCTGCTGCTAGCTAGTGAAGAGCGCTGACAGTCTACTCCTTTC 207  
DB 61 TCGGCGTGAAGACTACCTGCTGCTAGCTAGTGAAGAGCGCTGACAGTCTACTCCTTTC 120  
QY 208 ACTGACCTTGCTATCTTCGCAATAG 235  
DB 121 ACTGACCTTGCTATCTTCGCAATAG 148

RESULT 10  
AAC55315  
ID AAC55315 standard; DNA: 87 BP.  
XX

AC AAC55315;  
XX 05-FEB-2001 (first entry)  
DE Human activation-induced cytidine deaminase exon 1 SEQ ID NO:11.  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; anti-allergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Disgeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
OS Homo sapiens.  
PN WO200058480-A1.  
PD 05-OCT-2000.  
PE 28-MAR-2000; 2000WO-JP01918.  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX (NISR ) JAPAN TOBACCO INC.  
PA (HONJ/) HONJO T.  
PI Honjo T, Muramatsu M;  
PI WPI; 2000-611715/58.  
DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX Claim 18; Page 150; 174pp; Japanese.  
XX The present invention describes an activation-induced cytidine deaminase  
XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
XX has cytidine activity similar to APOBEC-1. AID has anti-allergic,  
XX antianemic, antiasthmatic, ophthalmological, anti-HIV and  
XX dermatological activities, and can be used in gene therapy. AID  
XX polynucleotides are useful in methods for identifying drugs for the  
XX treatment of B cell associated immune system disorders, immunodeficiency  
XX diseases and allergies, such as immunoglobulin A (IGA) deficiency  
XX disease, IGA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
XX colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
XX disease, Disgeorge disease, ataxia telangiectasia, common variable  
XX immunodeficiency disorder, MHC (major histocompatibility class) class  
XX II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
XX IgG disorder, and IgG subclass selection disorder. The DNA sequences  
XX encoding AID may be used for gene therapy and the antibodies to the AID  
XX protein may be used for diagnosis and treatment of these disorders. The  
XX present sequence represents the exon 1 genomic DNA sequence of human AID.  
SQ Sequence 87 BP; 28 A; 17 C; 23 G; 19 T; 0 other;  
Query Match 3.1%; Score 87; DB 21; Length 87;  
Best Local Similarity 100.0%; Pred. No. 2.2e+23;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGACCATCATTAATGAAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGACA 60  
DB 1 AGAGACCATCATTAATGAAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGACA 60  
QY 61 AGACACTCTGGACACCACTATGACAG 87  
DB 61 AGACACTCTGGACACCACTATGACAG 87

RESULT 11  
AAC55313  
ID AAC55313 standard; DNA: 5514 BP.  
XX  
AC AAC55313;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:9.  
XX  
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianaemic; antiastrumatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; Igg subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000WO-JP01918.  
XX  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
PA (NISEB) JAPAN TOBACCO INC.  
PA (HONJ/) HONJO T.  
PI Honjo T, Muramatsu M;  
PI WPI; 2000-611715/58.  
XX  
DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
PS Claim 17; Page 142-145; 174pp; Japanese.  
XX  
CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianaemic, antiastrumatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IGA) deficiency  
CC disease, IGA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, Digorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disease, MHC (major histocompatibility class)  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC Igg disorder, and Igg subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents a genomic DNA sequence of human AID.  
XX  
SQ Sequence 5514 BP; 1709 A; 1045 C; 1134 G; 1623 T; 3 other:  
Query Match 3.1%; Score 87; DB 21; Length 5514;  
Best Local Similarity 100.0%; Pred. No. 1.2e-23;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1032 AGAGACCATCATTAATGAGATTTTCTGCGCTGAGACTTGACGAGGACGACAGA 1091  
|||  
QY 61 AGACACTGTGACACCACTATGACAG 87  
|||||  
Db 1092 AGACACTGTGACACCACTATGACAG 1118  
|||||  
RESULT 12  
AAC55318  
ID AAC55318 standard; DNA: 116 BP.  
XX  
AC AAC55318;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase exon 4 SEQ ID NO:14.  
XX  
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianaemic; antiastrumatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; Igg subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000WO-JP01918.  
XX  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
PA (NISEB) JAPAN TOBACCO INC.  
PA (HONJ/) HONJO T.  
PI Honjo T, Muramatsu M;  
PI WPI; 2000-611715/58.  
XX  
DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
PS Claim 18; Page 151; 174pp; Japanese.  
XX  
CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianaemic, antiastrumatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IGA) deficiency  
CC disease, IGA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, Digorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disease, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC Igg disorder, and Igg subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents the exon 4 genomic DNA sequence of human AID.  
XX  
SQ Sequence 116 BP; 32 A; 25 C; 24 G; 35 T; 0 other;

Query Match 2.8%; Score 78; DB 21; Length 116;  
Best Local Similarity 100.0%; Pred. No. 5,6e-20;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 GAAAGACTTCAAGCGTGGGAGGCGTCATGAATTCAGTCTCTCCAGACG 604  
DB 39 GAAAGACTTCAAGCGTGGGAGGCGTCATGAATTCAGTCTCTCCAGACG 98  
QY 605 CTTCGGCGCATCCTTTTG 622  
DB 99 CTTCGGCGCATCCTTTTG 116

RESULT 13  
AAK76441/c  
ID AAK76441 standard; DNA; 280 BP.  
XX  
AC AAK76441;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31253.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW Cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN W0200157182-A2.  
PD  
XX 09-AUG-2001.  
PF 17-JAN-2001; 2001W0-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230457.  
PR 06-SEP-2000; 2000US-0230458.  
PR 06-SEP-2000; 2000US-0230439.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246529.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250361.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 05-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX  
XX  
XX Disclosure; SEQ ID NO 31253; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
XX represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 280 BP; 53 A; 88 C; 70 G; 69 T; 0 other.  
XX  
XX Query Match 2.6%; Score 74; DB 22; Length 280;  
XX Best Local Similarity 100.0%; Pred. No. 1.6e-18;  
XX Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAATCCCACTCTGGAGGCTGAGGAGAGAAATCGTTGAACCCAGAG 2041  
DB 126 GGCACCTGTAATCCCACTCTGGAGGCTGAGGAGAGAAATCGTTGAACCCAGAG 67  
QY 2042 GTGAGGTTGCAGT 2055  
DB 66 GTGAGGTTGCAGT 53

RESULT 14  
ABV22611/c

ID ABV22611 standard; cDNA; 481 BP.  
XX  
XX AC ABV22611;  
XX  
XX 13-SEP-2002 (first entry)  
XX  
XX DE Human prostate expression marker cDNA 22602.  
XX  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200160860-A2.  
XX  
XX PD 23-AUG-2001.  
XX  
XX PE 20-FEB-2001; 2001WO-US05171.  
XX  
XX PR 17-FEB-2000; 2000US-183319P.  
XX PR 16-MAR-2000; 2000US-189862P.  
XX PR 25-MAY-2000; 2000US-207454P.  
XX PR 09-JUN-2000; 2000US-211314P.  
XX PR 18-JUL-2000; 2000US-219007P.  
XX PR 13-DEC-2000; 2000US-255281P.  
XX  
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX PI Schlegel R, Endege WO, Monahan JE;  
XX  
XX DR WPI: 2001-662795/76.  
XX  
XX XX  
XX PT Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX  
XX PS Claim 1; Page 3959; 11750pp; English.  
XX  
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX specification or its complement. (I) is useful for:  
XX (a) assessing whether a patient is afflicted with prostate cancer;  
XX (b) monitoring the progression of prostate cancer in a patient;  
XX (c) assessing the efficacy of a test compound to inhibit prostate  
XX cancer in a patient;  
XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
XX in a patient;  
XX (e) selecting a composition for inhibiting prostate cancer in a patient;  
XX (f) assessing the prostate cell carcinogenic potential of a compound;  
XX (g) determining whether prostate cancer has metastasized in a patient;  
XX (h) assessing the aggressiveness or indolence of prostate cancer in a  
XX patient;  
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX Sequence 481 BP; 111 A; 103 C; 111 G; 153 T; 3 other;  
XX  
XX Query Match 2.5%; Score 70; DB 23; Length 481;  
XX Best Local Similarity 100.0%; Pred. No. 4.8e-17;  
XX Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCCACTCTGGAGGCTGAGGAGAGAAATCGTTGAACCCAGAGGTG 2045  
DB 204 CCTGTAAATCCCACTCTGGAGGCTGAGGAGAGAAATCGTTGAACCCAGAGGTG 145  
QY 2046 AGGTTCAGT 2055  
DB 144 AGGTTCAGT 135

RESULT 15  
ABV28433/c  
ID ABV28433 standard; cDNA; 481 BP.  
XX



```

AC ABV28433;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker CDNA 28424.
XX
KW Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene, ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 5931; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 481 BP; 111 A; 103 C; 111 G; 153 T; 3 other;

Query Match 2.5%; Score 70; DB 23; Length 481;
Best Local Similarity 100.0%; Pred. No. 4.8e-17;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCGTGAATCCGAGCTCTGGAGGCTGAGCAGAGAAATCGTTGAACCCAGAGGTGG 2045
DB 204 CCGTGAATCCGAGCTCTGGAGGCTGAGCAGAGAAATCGTTGAACCCAGAGGTGG 145
QY 2046 AGTTGCAGT 2055
DB 144 AGTTGCAGT 135

```

Search completed: July 7, 2003, 20:19:18  
 Job time : 613 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: July 7, 2003, 20:11:31 ; Search time 156 Seconds  
(Without alignments)  
5539.840 Million cell updates/sec

Title: US-09-966-880a-7  
Perfect score: 2818  
Sequence: 1 agagaaccatcataatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: Issued\_Patents\_NA:\*  
2: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfileseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
C 1	59	2.1	99500	4 US-09-798-096-10	Sequence 10, Appl
C 2	57	2.0	84495	4 US-09-797-906-3	Sequence 3, Appl
C 3	57	2.0	111282	4 US-09-754-250-3	Sequence 3, Appl
C 4	57	2.0	162450	4 US-09-345-882-1	Sequence 1, Appl
C 5	55	2.0	1278	2 US-08-909-965C-4	Sequence 4, Appl
C 6	53	1.9	176373	3 US-09-128-155-17	Sequence 17, Appl
C 7	52	1.8	341	4 US-09-404-879A-136	Sequence 136, App
C 8	52	1.8	451	4 US-09-404-879A-1	Sequence 1, Appl
C 9	52	1.8	461	4 US-09-404-879A-3	Sequence 3, Appl
C 10	52	1.8	7152	4 US-09-167-681-29	Sequence 29, Appl
C 11	52	1.8	87350	3 US-08-781-891-79	Sequence 79, Appl
C 12	52	1.8	87543	4 US-09-791-211-3	Sequence 3, Appl
C 13	51	1.8	1701	4 US-09-078-294-9	Sequence 9, Appl
C 14	51	1.8	2174	4 US-09-613-444-1	Sequence 1, Appl
C 15	51	1.8	3844	4 US-09-689-423-1	Sequence 1, Appl
C 16	51	1.8	8396	4 US-09-328-174A-1	Sequence 1, Appl
C 17	51	1.8	8409	4 US-09-167-681-37	Sequence 37, Appl
C 18	51	1.8	36159	4 US-09-749-588-3	Sequence 3, Appl
C 19	51	1.8	38564	4 US-09-734-673-3	Sequence 3, Appl
C 20	50	1.8	1260	1 US-08-599-252-83	Sequence 83, Appl
C 21	50	1.8	1260	1 US-08-436-074-56	Sequence 56, Appl
C 22	50	1.8	1260	1 PCT-US96-06352-83	Sequence 83, Appl
C 23	50	1.8	1260	5 PCT-US96-06583-83	Sequence 83, Appl
C 24	50	1.8	1442	2 US-08-454-557C-120	Sequence 120, App
C 25	50	1.8	1442	2 US-08-340-425D-120	Sequence 120, App
C 26	50	1.8	1442	2 US-08-450-673C-120	Sequence 120, App
C 27	50	1.8	12565	4 US-09-345-217-3	Sequence 3, Appl

C 28	50	1.8	29629	4 US-09-729-995-3	Sequence 3, Appl
C 29	50	1.8	36651	4 US-09-738-894A-3	Sequence 3, Appl
C 30	50	1.8	40000	4 US-09-780-049-18	Sequence 18, Appl
C 31	50	1.8	45546	4 US-09-146-053-6	Sequence 6, Appl
C 32	49	1.7	201	2 US-08-849-701-5	Sequence 5, Appl
C 33	49	1.7	35060	3 US-08-814-095-7	Sequence 7, Appl
C 34	49	1.7	43950	4 US-09-735-934A-3	Sequence 3, Appl
C 35	49	1.7	168575	4 US-09-426-290-1	Sequence 1, Appl
C 36	48	1.7	985	4 US-09-641-628-350	Sequence 350, App
C 37	48	1.7	6769	1 US-08-480-784-20	Sequence 20, Appl
C 38	48	1.7	6769	1 US-08-483-553-20	Sequence 20, Appl
C 39	48	1.7	6769	1 US-08-487-002-20	Sequence 20, Appl
C 40	48	1.7	6769	1 US-08-483-554B-20	Sequence 20, Appl
C 41	48	1.7	6769	1 US-08-488-011B-20	Sequence 20, Appl
C 42	48	1.7	6769	4 US-08-850-727-20	Sequence 20, Appl
C 43	48	1.7	6769	5 PCT-US95-10202-20	Sequence 20, Appl
C 44	48	1.7	6769	5 PCT-US95-10203-20	Sequence 20, Appl
C 45	48	1.7	6769	5 PCT-US95-10220-20	Sequence 20, Appl

#### ALIGNMENTS

RESULT 1  
US-09-798-096-10/C  
Sequence 10, Application US/09798096  
Patent No. 639378  
GENERAL INFORMATION:  
APPLICANT: Donna T. Ward  
TITLE OF INVENTION: ANTISENSE MODULATION OF RECOIL2 EXPRESSION  
FILE REFERENCE: RTS 0207  
CURRENT APPLICATION NUMBER: US/09/798,096  
CURRENT FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 89  
SEQ ID NO 10  
LENGTH: 99500  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
US-09-798-096-10

Query Match 2.1%; Score 59; DB 4; Length 99500;  
Best Local Similarity 100.0%; Pred. No. 1.5e-14;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TCCGAGCTACTGGAGGCGTGAGGAGAGATCGCTGAACCCGAGAGGTGG 2051  
DB 57976 TCCGAGCTACTGGAGGCGTGAGGAGAGATCGCTGAACCCGAGAGGTGG 57918

RESULT 2  
US-09-797-906-3  
Sequence 3, Application US/09797906  
Patent No. 6329188  
GENERAL INFORMATION:  
APPLICANT: Zhanqie YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY  
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
FILE REFERENCE: CLO01151CIP  
CURRENT APPLICATION NUMBER: US/09/797,906  
CURRENT FILING DATE: 2001-03-05  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 84495  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(84495)  
OTHER INFORMATION: n = A,T,C or G

US-09-797-906-3

Query Match 2.0%; Score 57; DB 4; Length 84495;  
Best Local Similarity 100.0%; Pred. No. 9.5e-14;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGACCCCTGTAATCCGACTACTGTGGAGGCTGAGGAGAGAAATCGCTTGACCC 2035  
DB 20037 GCAGGACCCCTGTAATCCGACTACTGTGGAGGCTGAGGAGAGAAATCGCTTGACCC 20093

RESULT 3

US-09-754-250-3/C  
Sequence 3, Application US/09754250  
Patent No. 6376225  
GENERAL INFORMATION:  
APPLICANT: WEI, Ming-Hui et al  
TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE  
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
FILE REFERENCE: C1001063  
CURRENT APPLICATION NUMBER: US/09/754,250  
CURRENT FILING DATE: 2001-01-05  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 111282  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(11282)  
OTHER INFORMATION: n = A,T,C or G  
US-09-754-250-3

Query Match 2.0%; Score 57; DB 4; Length 111282;  
Best Local Similarity 100.0%; Pred. No. 9.1e-14;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTATCCGACTACTGTGGAGGCTGAGGAGAGATCCGCTTAACCCAGGAGG 2042  
DB 106427 CCTGTATCCGACTACTGTGGAGGCTGAGGAGAGATCCGCTTAACCCAGGAGG 106371

RESULT 4

US-09-345-882-1  
Sequence 1, Application US/09345882  
Patent No. 6393773  
GENERAL INFORMATION:  
APPLICANT: Bougueteloret, Lydie  
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)  
TITLE OF INVENTION: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.  
FILE REFERENCE: GENSET.031A  
CURRENT APPLICATION NUMBER: US/09/345,882  
CURRENT FILING DATE: 1999-06-30  
PRIOR APPLICATION NUMBER: US 60/091,315  
PRIOR FILING DATE: 1998-06-30  
PRIOR APPLICATION NUMBER: US 60/111,909  
PRIOR FILING DATE: 1998-12-10  
NUMBER OF SEQ ID NOS: 140  
SOFTWARE: Patent.pm  
SEQ ID NO 1  
LENGTH: 162450  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 72794  
OTHER INFORMATION: 5-124-273 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 88073  
OTHER INFORMATION: 5-127-261 : polymorphic base A or C

FEATURE:  
NAME/KEY: allele  
LOCATION: 90842  
OTHER INFORMATION: 99-1437-325 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 93714  
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97122  
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97152  
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99098  
OTHER INFORMATION: 5-130-257 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99117  
OTHER INFORMATION: 5-130-276 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 103806  
OTHER INFORMATION: 5-131-395 : polymorphic base A or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106940  
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108106  
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108149  
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108471  
OTHER INFORMATION: 5-136-174 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 134134  
OTHER INFORMATION: 5-140-120 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 134362  
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A  
FEATURE:  
NAME/KEY: allele  
LOCATION: 134374  
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA  
FEATURE:  
NAME/KEY: allele  
LOCATION: 146328  
OTHER INFORMATION: 5-143-84 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 146345  
OTHER INFORMATION: 5-143-101 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 150329  
OTHER INFORMATION: 5-145-24 : polymorphic base A or G  
FEATURE:

NAME/KEY: allele  
LOCATION: 160031  
OTHER INFORMATION: 5-148-352 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 72771..72817  
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30  
FEATURE:  
NAME/KEY: allele  
LOCATION: 72771..72817  
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51  
FEATURE:  
NAME/KEY: allele  
LOCATION: 88050..88096  
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31  
FEATURE:  
NAME/KEY: allele  
LOCATION: 88050..88096  
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52  
FEATURE:  
NAME/KEY: allele  
LOCATION: 90819..90865  
OTHER INFORMATION: complement polymorphic fragment 9-1437-325 SEQ ID49  
FEATURE:  
NAME/KEY: allele  
LOCATION: 90819..90865  
OTHER INFORMATION: complement polymorphic fragment 9-1437-325 SEQ ID70  
FEATURE:  
NAME/KEY: allele  
LOCATION: 93690..93736  
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32  
FEATURE:  
NAME/KEY: allele  
LOCATION: 93690..93736  
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97099..97145  
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97099..97145  
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97130..97177  
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97130..97177  
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99075..99121  
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99075..99121  
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99094..99140  
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99094..99140  
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56  
FEATURE:  
NAME/KEY: allele  
LOCATION: 103783..103828  
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36  
FEATURE:  
NAME/KEY: allele

LOCATION: 103783..103828  
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106918..106966  
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106918..106966  
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108084..108130  
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108084..108130  
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108127..108177  
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108127..108177  
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60  
FEATURE:  
Query Match 2.0%; Score 57; DB 4; Length 162450;  
Best Local Similarity 100.0%; Pred. No. 8.6e-14;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DY 1986 CCTGTAATCCAGCTACTTGAGGCTGAGCGAGGAATGCTGAAACCCAGGAGG 2042  
DB 136970 CCTGTAATCCAGCTACTTGAGGCTGAGCGAGGAATGCTGAAACCCAGGAGG 137026  
RESULT 5  
US-08-909-965C-4  
Sequence 4, Application US/08909965C  
Patent No. 5936078  
GENERAL INFORMATION:  
APPLICANT: Kuga Tetsuo  
APPLICANT: Nakagawa Satoshi  
APPLICANT: Sakaki Yoshiyuki  
APPLICANT: Zhao Nanding  
APPLICANT: Hashida Hideji  
TITLE OF INVENTION: NOVEL DNA, NOVEL POLYPEPTIDE  
TITLE OF INVENTION: AND NOVEL ANTIBODY  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FITZPATRICK, CELLA, HARPER AND SCINTO  
STREET: 277 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10172-0194  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/909,965C  
FILING DATE: August 12, 1997  
CLASSIFICATION: 514  
PRIOR APPLICATION NUMBER: JP 322745/95  
APPLICATION NUMBER: PCT/JP96/03630  
FILING DATE: 12-NO. 5936078-1995  
FILING DATE: 12-Dec-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Lawrence S. Perry

REGISTRATION NUMBER: 31865  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-2400  
TELEFAX: 212-758-2982  
TELEX: 236262  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1278 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORGANISM: human  
IMMEDIATE SOURCE:  
CLONE: F328  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 166 to 261  
LOCATION: 704 to 829  
IDENTIFICATION METHOD: by experiment  
US-08-909-965c-4

Query Match 2.0%; Score 55; DB 2; Length 1278;  
Best Local Similarity 100.0%; Pred. No. 1.1e-12;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1987 CTGTATCCAGTACTGTGGAGCGTGGAGGAGAGAACTGTTAACCAGAG 2041  
DB 1129 CTGTATCCAGTACTGTGGAGCGTGGAGGAGAGAACTGTTAACCAGAG 1183

RESULT 6  
US-09-128-155-17  
Sequence 17 Application US/09128155  
Patent No. 6117654  
GENERAL INFORMATION:  
APPLICANT: Pan, Yang  
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
FILE REFERENCE: 09404/052001  
CURRENT APPLICATION NUMBER: US/09/128,155  
CURRENT FILING DATE: 1998-08-03  
EARLIER APPLICATION NUMBER: US 60/091,650  
EARLIER FILING DATE: 1998-07-02  
EARLIER APPLICATION NUMBER: US 60/054,646  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 17  
LENGTH: 176373  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(176373)  
OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-17

Query Match 1.9%; Score 53; DB 3; Length 176373;  
Best Local Similarity 100.0%; Pred. No. 3.1e-12;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1990 TATCCAGTACTGTGGAGCGTGGAGGAGAGAACTGTTAACCAGAG 2042  
DB 98392 TATCCAGTACTGTGGAGCGTGGAGGAGAGAACTGTTAACCAGAG 98444

RESULT 7  
US-09-404-879A-136/c  
Sequence 136 Application US/09404879A  
Patent No. 6468546  
GENERAL INFORMATION:

APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.  
APPLICANT: Algate, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.46202  
CURRENT APPLICATION NUMBER: US/09/404,879A  
CURRENT FILING DATE: 1999-09-24  
NUMBER OF SEQ ID NOS: 393  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 136  
LENGTH: 341  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-404-879A-136

Query Match 1.8%; Score 52; DB 4; Length 341;  
Best Local Similarity 100.0%; Pred. No. 1.9e-11;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1829 GGGCCGGGCGTGTGGCTACGCTGTATCCAGCACTTGGAGGCCGAG 1880  
DB 114 GGGCCGGGCGTGTGGCTACGCTGTATCCAGCACTTGGAGGCCGAG 63

RESULT 8  
US-09-404-879A-1/c  
Sequence 1 Application US/09404879A  
Patent No. 6468546  
GENERAL INFORMATION:  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.  
APPLICANT: Algate, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.46202  
CURRENT APPLICATION NUMBER: US/09/404,879A  
CURRENT FILING DATE: 1999-09-24  
NUMBER OF SEQ ID NOS: 393  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 461  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-404-879A-1

Query Match 1.8%; Score 52; DB 4; Length 461;  
Best Local Similarity 100.0%; Pred. No. 1.8e-11;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1829 GGGCCGGGCGTGTGGCTACGCTGTATCCAGCACTTGGAGGCCGAG 1880  
DB 364 GGGCCGGGCGTGTGGCTACGCTGTATCCAGCACTTGGAGGCCGAG 313

RESULT 9  
US-09-404-879A-3/c  
Sequence 3 Application US/09404879A  
Patent No. 6468546  
GENERAL INFORMATION:  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.  
APPLICANT: Algate, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.46202  
CURRENT APPLICATION NUMBER: US/09/404,879A  
CURRENT FILING DATE: 1999-09-24  
NUMBER OF SEQ ID NOS: 393  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 461  
TYPE: DNA

ORGANISM: Homo sapiens  
US-09-404-879A-3

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.8e-11;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1829 GGGCGGGCGTGGTGGCTACAGCCTGTATCCAGCACTTGGAGCGCGAG 1880  
DB 364 GGGCGGGCGTGGTGGCTACAGCCTGTATCCAGCACTTGGAGCGCGAG 313

RESULT 10  
US-09-167-681-29/c  
Sequence 29, Application US/09167681A  
Patent No. 6265561  
GENERAL INFORMATION:  
APPLICANT: Weinsztein, M.D., Richard M.  
APPLICANT: Rafilogianis, Rebecca B.  
APPLICANT: Wood, Thomas C.  
APPLICANT: O'Brien, Diane M.  
TITLE OF INVENTION: SULFOTRANSFERASE SEQUENCE VARIANTS  
FILE REFERENCE: 07039/118001  
CURRENT APPLICATION NUMBER: US/09/167,681A  
CURRENT FILING DATE: 1998-10-07  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 29  
LENGTH: 7152  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3810)...(3956)  
NAME/KEY: CDS  
LOCATION: (4061)...(4186)  
NAME/KEY: CDS  
LOCATION: (4276)...(4374)  
NAME/KEY: CDS  
LOCATION: (5584)...(5709)  
NAME/KEY: CDS  
LOCATION: (5805)...(5900)  
NAME/KEY: CDS  
LOCATION: (6426)...(6605)  
NAME/KEY: CDS  
LOCATION: (6728)...(6837)  
US-09-167-681-29

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.2e-11;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTGGAGGCTGAGCGAGAGATCGCTGAGCCAGGAG 2042  
DB 774 AATCCAGCTACTGGAGGCTGAGCGAGAGATCGCTGAGCCAGGAG 723

RESULT 11  
US-08-781-891-79/c  
Sequence 79, Application US/08781891  
Patent No. 6090620  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-Sh  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg, Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-79

Query Match  
Best Local Similarity 100.0%; Pred. No. 8.3e-12;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGACCTGTATCCAGCTTCTGGAGGCTGAGCGAGAGATCGCTT 2030  
DB 78361 GCAGGACCTGTATCCAGCTTCTGGAGGCTGAGCGAGAGATCGCTT 78480

RESULT 12  
US-09-791-211-3/c  
Sequence 3, Application US/09791211  
Patent No. 6448080  
GENERAL INFORMATION:  
APPLICANT: Donna T. Ward  
TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION  
FILE REFERENCE: RTS-0205  
CURRENT APPLICATION NUMBER: US/09/791,211  
CURRENT FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 90  
SEQ ID NO 3  
LENGTH: 87543  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: 7421  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 7427  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 11609  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 12605  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 12742  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29370  
OTHER INFORMATION: unknown  
NAME/KEY: unsure

LOCATION: 29422  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29979  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29980  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29981  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 30136  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 30140  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 31205  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 31206  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 31582  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 33095  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 33160  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 34066  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 34072  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 36816  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 39020  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 42164  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 42459  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 46808  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 46823  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 46826  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 47291  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 52786  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 52787  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 53384  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 54684

OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 59215  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 59235  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 59242  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 63290  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 66614  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68660  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68697  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68718  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68733  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68739  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 69785  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 79134  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 79198  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 86336  
OTHER INFORMATION: unknown  
OTHER INFORMATION: unknown  
US-09-791-211-3

Query Match 1.8%; Score 52; DB 4; Length 87543;  
Best Local Similarity 100.0%; Pred. No. 8.3e-12;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGACGAGAAATCGCTT 2030  
DB 78724 GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGACGAGAAATCGCTT 78673

RESULT 13  
US-09-078-294-9/c  
Sequence 9, Application US/09078294  
Patent No. 6265211  
GENERAL INFORMATION:  
APPLICANT: Choo, Kong-Hong Andy  
APPLICANT: Du Sart, Desiree  
APPLICANT: Cancilla, Michael R.  
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
FILE REFERENCE: Davies Col  
CURRENT APPLICATION NUMBER: US/09/078,294  
CURRENT FILING DATE: 1998-05-13  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn ver. 2.0  
SEQ ID NO 9  
LENGTH: 1701  
TYPE: DNA  
ORGANISM: BAC-F2 contlg 5

US-09-078-294-9

## Query Match

1.8%; Score 51; DB 4; Length 1701;  
Best Local Similarity 100.0%; Pred. No. 3.7e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGAGATCGCTTGAACCC 2036  
|||||  
DB 1246 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGAGATCGCTTGAACCC 1196

## RESULT 14

US-09-613-444-1  
; Sequence 1, Application US/09613444  
; Patent No. 6444427  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig, Erwin H.  
; APPLICANT: Farese, Robert V.  
; APPLICANT: Innerarity, Thomas L.  
; APPLICANT: Cases, Sylvaine  
; TITLE OF INVENTION: Polymorphisms in a Diacylglycerol  
; FILE REFERENCE: 6510191051  
; CURRENT APPLICATION NUMBER: US/09/613,444  
; CURRENT FILING DATE: 2000-07-11  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2174  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-613-444-1

## Query Match

1.8%; Score 51; DB 4; Length 2174;  
Best Local Similarity 100.0%; Pred. No. 3.5e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2005 GGGAGGCTGAGGAGAGATCGCTTGAACCCAGAGGCTGAGGCTTGCAGT 2055  
|||||  
DB 454 GGGAGGCTGAGGAGAGATCGCTTGAACCCAGAGGCTGAGGCTTGCAGT 504

## RESULT 15

US-09-689-423-1/c  
; Sequence 1, Application US/09689423  
; Patent No. 6414131  
; GENERAL INFORMATION:  
; APPLICANT: Berrettini, Wade H.  
; TITLE OF INVENTION: Gene and Methods for Diagnosing Neuropsychiatric  
; FILE REFERENCE: PENN-0731  
; CURRENT APPLICATION NUMBER: US/09/689,423  
; CURRENT FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: 60/195,620  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: 60/159,354  
; PRIOR FILING DATE: 1999-10-14  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 3844  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-689-423-1

## Query Match

1.8%; Score 51; DB 4; Length 3844;  
Best Local Similarity 100.0%; Pred. No. 3.2e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGAGATCGCTTGAACCC 2036  
|||||  
DB 1536 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGAGATCGCTTGAACCC 1486

Search completed: July 7, 2003, 23:19:41  
Job time: 160 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 7, 2003, 20:19:21 ; Search time 489 Seconds  
(without alignments)  
8995.652 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 agagAACCATCAATTAATGA.....AAAAAAAAAAAAAAAAAAAA 2818

Scoring table:  
OLIGO\_NUC  
Gap60 60.0 , Gapext 60.0

Searched: 1085931 seqs, 780495707 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCF\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCFUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2818	100.0	2818	9 US-09-966-880A-7	Sequence 7, Appl1
2	2174	77.1	6564	9 US-09-966-880A-10	Sequence 10, Appl1
3	2174	77.1	11204	9 US-09-966-880A-35	Sequence 35, Appl1
4	2172	77.1	2172	9 US-09-966-880A-15	Sequence 15, Appl1
5	271	9.6	271	9 US-09-966-880A-13	Sequence 13, Appl1
6	148	5.3	148	9 US-09-966-880A-12	Sequence 12, Appl1
7	87	3.1	87	9 US-09-966-880A-11	Sequence 11, Appl1
8	87	3.1	5514	9 US-09-966-880A-9	Sequence 9, Appl1
9	78	2.8	116	9 US-09-966-880A-14	Sequence 14, Appl1
10	70	2.5	167343	10 US-09-962-436-281	Sequence 281, Appl1
11	70	2.5	167343	10 US-09-964-824A-273	Sequence 273, Appl1
12	66	2.3	125	9 US-10-074-095-922	Sequence 922, Appl1
13	66	2.3	125	9 US-10-074-095-922	Sequence 922, Appl1
14	66	2.3	125	10 US-09-764-860-922	Sequence 922, Appl1
15	66	2.3	125	10 US-09-764-860-922	Sequence 923, Appl1
16	66	2.3	451	10 US-09-867-701-10468	Sequence 10468, A
17	66	2.3	488	9 US-09-918-995-1236	Sequence 1236, Ap
18	64	2.3	17216	10 US-09-764-877-3565	Sequence 3565, Ap
19	64	2.3	17217	10 US-09-764-877-3566	Sequence 3566, Ap

C 20	64	2.3	32249	9 US-09-764-891-7619	Sequence 7619, Ap
C 21	63	2.3	139257	9 US-09-920-671-111	Sequence 11, Appl1
C 22	64	2.2	31730	10 US-09-764-877-3810	Sequence 3810, Ap
C 23	61	2.2	15275	9 US-10-091-504-1475	Sequence 1475, Ap
C 24	61	2.2	15275	10 US-09-764-869-1475	Sequence 1475, Ap
C 25	60	2.1	118	9 US-10-072-349-295	Sequence 295, Appl1
C 26	60	2.1	118	10 US-09-764-855-295	Sequence 295, Appl1
C 27	60	2.1	6427	9 US-10-072-349-292	Sequence 292, Appl1
C 28	60	2.1	6427	10 US-09-764-855-292	Sequence 292, Appl1
C 29	60	2.1	11173	9 US-09-898-556A-10	Sequence 3, Appl1
C 30	59	2.1	83450	9 US-09-811-469-3	Sequence 1543, Ap
C 31	58	2.1	55795	10 US-09-880-107-1543	Sequence 1975, Ap
C 32	57	2.0	129	9 US-10-092-154-1975	Sequence 1975, Ap
C 33	57	2.0	129	10 US-09-764-847-1975	Sequence 1975, Ap
C 34	57	2.0	157	9 US-09-764-891-6037	Sequence 6037, Ap
C 35	57	2.0	157	10 US-09-764-877-2653	Sequence 2653, Ap
C 36	57	2.0	445	9 US-09-918-995-25881	Sequence 25881, A
C 37	57	2.0	517	9 US-10-060-036-3361	Sequence 3361, Ap
C 38	57	2.0	1382	9 US-10-092-154-1253	Sequence 1253, Ap
C 39	57	2.0	1382	10 US-09-764-847-1253	Sequence 1253, Ap
C 40	57	2.0	5670	10 US-09-954-456-1146	Sequence 1146, Ap
C 41	57	2.0	10812	9 US-09-764-891-8995	Sequence 8995, Ap
C 42	57	2.0	15261	9 US-09-764-891-10185	Sequence 10185, A
C 43	57	2.0	17286	10 US-09-764-877-3234	Sequence 3234, Ap
C 44	57	2.0	28770	10 US-09-817-198A-3	Sequence 3, Appl1
C 45	57	2.0	32216	9 US-09-764-891-9613	Sequence 9613, Ap

## ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
Sequence 7, Application US/09966880A  
Patent No. US20020164742A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JPO0/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178939  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (80)...(673)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(79)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (677)...(2818)  
US-09-966-880A-7

Query Match 100.0%; Score 2818; DB 9; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 AGAGAACCATCAATTAATGAAGTGAAGATTTCGCGCTGAGACTTGACGAGGACAGAGA 60  
|||||

D	b	1	AGAGAACATCATTAATTAATGAGTGAAGATTTTTCCTGCTAGACTTGCAGGAGACCAAGA	60
O	y	61	AGACACTCTGGACACACACTATGAGACGCCCTTGATGTAACCGAGGAAGATTCTTTTACCA	120
D	b	61	AGACACTCTGGACACACACTATGAGAGGCCCTTGATGTAACCGAGGAAGATTCTTTACCA	120
O	y	121	ATTCAAAAATTCCTCGCTGGGCTAAGGGTGGCGGTGAGACCTTACCTGTGCTACGTAGTGA	180
D	b	121	ATTCAAAAATTCCTCGCTGGGCTAAGGGTGGCGGTGAGACCTTACCTGTGCTACGTAGTGA	180
O	y	181	GAGCGGTACAGTGTATACCTTTTTCACGTGGACTTGGTATATCTGCACATTAAGAACGG	240
D	b	181	GAGCGGTACAGTGTATACCTTTTTCACGTGGACTTGGTATATCTGCACATTAAGAACGG	240
O	y	241	CTGCCACGTGAATTTGCTCTCTCCGCTACATCTCGGACTGAGGACCTAGACCTGAGCGG	300
D	b	241	CTGCCACGTGAATTTGCTCTCTCCGCTACATCTCGGACTGAGGACCTAGACCTGAGCGG	300
O	y	301	CTGCTACCGGGTCACTGTGTTCACTCTCTGAGGCCCTGTAACGACTGTGCCGACATGT	360
D	b	301	CTGCTACCGGGTCACTGTGTTCACTCTCTGAGGCCCTGTAACGACTGTGCCGACATGT	360
O	y	361	GGCGGACTTCTGGGAGGGAACCCCAACTAGTGTGAGATCTTACACGGGGCCCTCTA	420
D	b	361	GGCGGACTTCTGGGAGGGAACCCCAACTAGTGTGAGATCTTACACGGGGCCCTCTA	420
O	y	421	CTTCTGTAGAGACGGCAAGCGCTGAGCCCGAGGGGCTGGGCGGTGACCCGCGGGGT	480
D	b	421	CTTCTGTAGAGACGGCAAGCGCTGAGCCCGAGGGGCTGGGCGGTGACCCGCGGGGT	480
O	y	481	GCAAAATAGCCATCATGACCTTCAAGATTATTTTACTGTGGAATACTTTTGTAGAAA	540
D	b	481	GCAAAATAGCCATCATGACCTTCAAGATTATTTTACTGTGGAATACTTTTGTAGAAA	540
O	y	541	CCATGAAGAAGACTTTCAAGCGCTGGGAGGGCTGTGATGAATAATCAGTCTCTCTCCAG	600
D	b	541	CCATGAAGAAGACTTTCAAGCGCTGGGAGGGCTGTGATGAATAATCAGTCTCTCTCCAG	600
O	y	601	ACAGCTTGGGGGACCTCTTTGCCCTGTATGAGAGTATGATACCTACAGACGATTTGCG	660
D	b	601	ACAGCTTGGGGGACCTCTTTGCCCTGTATGAGAGTATGATACCTACAGACGATTTGCG	660
O	y	661	TACTTTGGGACTTTGATAGACAACTTCCAGGATGTACACACGATGAATATCTCTGCTG	720
D	b	661	TACTTTGGGACTTTGATAGACAACTTCCAGGATGTACACACGATGAATATCTCTGCTG	720
O	y	721	AAGACAGTGGATATAAAAAACACTCCTTCAAGCTTCTCTGTTTATATCTTAACCTTCAC	780
D	b	721	AAGACAGTGGATATAAAAAACACTCCTTCAAGCTTCTCTGTTTATATCTTAACCTTCAC	780
O	y	781	TTTCTTAGAGTTTACAGAAAAAATATTATATATAGACTCTTTAAAAAGATCTATGCTGTG	840
D	b	781	TTTCTTAGAGTTTACAGAAAAAATATTATATATAGACTCTTTAAAAAGATCTATGCTGTG	840
O	y	841	AAATATAGGAAGAACACAGGTCGTGGCCAGGAGCGTGTCAATTTGTCAGTTTGAAT	900
D	b	841	AAATATAGGAAGAACACAGGTCGTGGCCAGGAGCGTGTCAATTTGTCAGTTTGAAT	900
O	y	901	GCAACATTTGCCCTACTGCGGAATATACAGAACTGCAAGACCTGGGAGATCCTTAAAGT	960
D	b	901	GCAACATTTGCCCTACTGCGGAATATACAGAACTGCAAGACCTGGGAGATCCTTAAAGT	960
O	y	961	CAAGCTTTTCTATAGACTTTTAGGTAGAGATGAGACAGAAAGTATCTTAAAAAGCATG	1020
D	b	961	CAAGCTTTTCTATAGACTTTTAGGTAGAGATGAGACAGAAAGTATCTTAAAAAGCATG	1020
O	y	1021	GTGAGAGGATCAAAATGTTTATATCAACATCCTTATTTATTTGATTCATTTGAGTTAAC	1080
D	b	1021	GTGAGAGGATCAAAATGTTTATATCAACATCCTTATTTATTTGATTCATTTGAGTTAAC	1080
O	y	1081	AGTGTGTATGATAGATTTTCTATTTCTTTCCCTTGACGTTTACTTCTTAAGTAAACAC	1140
D	b	1081	AGTGTGTATGATAGATTTTCTATTTCTTTCCCTTGACGTTTACTTCTTAAGTAAACAC	1140

QY		1141	AAACTCTTCCATCAGGCCGACGATCTATAGAGACCTCCTAATAGAGATCTGGGGATGT	1200
Dd		1141	AAACTCTTCATCGAGGCATCATCTATAGGACCCTCAATAGAGATCTGGGGATGT	1200
QY		1201	GACCCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGGTGATGTTTAAACAG	1260
Dd		1201	GACCCCAACCATCTCTCCAAAGCATTAATATCCAATCATGCGGTGATGTTTAAACAG	1260
QY		1261	CAGAAGCATGTTTTATGTTTGTCAAAAGAGATTGTTATGGGTGGGATGAGGTATA	1320
Dd		1261	CAGAAGCATGTTTTATGTTTGTCAAAAGAGATTGTTATGGGTGGGATGAGGTATA	1320
QY		1321	GACCATGATGCTACCTTCAAGTACTTTATPAAGAGATCTTAAATGGGCGAGAGAC	1380
Dd		1321	GACCATGATGCTACCTTCAAGTACTTTATPAAGAGATCTTAAATGGGCGAGAGAC	1380
QY		1381	TGTGAACAAGACACCCCTAATATATGSGTGTGTTGAAGTAGCAAACTCTTGAAACGC	1440
Dd		1381	TGTGAACAAGACACCCCTAATATATGSGTGTGTTGAAGTAGCAAACTCTTGAAACGC	1440
QY		1441	AAACTCTTTTAGAGNAGTCCTAATTTAGAAACACCCACAACCTTCACATCATCATTA	1500
Dd		1441	AAACTCTTTTAGAGNAGTCCTAATTTAGAAACACCCACAACCTTCACATCATCATTA	1500
QY		1501	GCAAAACAATTGGAAGAAAGTTGCTTGAATGTTGGGAGAGAAAAATCTATTGGCTCGT	1560
Dd		1501	GCAAAACAATTGGAAGAAAGTTGCTTGAATGTTGGGAGAGAAAAATCTATTGGCTCGT	1560
QY		1561	GGGTCTCTCATCTCACAAAATGCCAATCAGSTGCAAGTTTGCTCATTTTGTATGTGT	1620
Dd		1561	GGGTCTCTCATCTCACAAAATGCCAATCAGSTGCAAGTTTGCTCATTTTGTATGTGT	1620
QY		1621	GATGCTTCTCCCAAGGTATATTAATTAATTAAGAGTTGTGCAAAACAGATGATTA	1680
Dd		1621	GATGCTTCTCCCAAGGTATATTAATTAATTAAGAGTTGTGCAAAACAGATGATTA	1680
QY		1661	AGTCGCGAACCCTGGCACACGCTCATATGTTCTAGCTCTTGGGAGTTGAGGAGGAGA	1740
Dd		1661	AGTCGCGAACCCTGGCACACGCTCATATGTTCTAGCTCTTGGGAGTTGAGGAGGAGA	1740
QY		1741	TGCGTTGAACACAGSTGTTCAAGCCGACCTGGGCAACATPAACAAGTCTGTCTTCAA	1800
Dd		1741	TGCGTTGAACACAGSTGTTCAAGCCGACCTGGGCAACATPAACAAGTCTGTCTTCAA	1800
QY		1801	AAAAAAAAAAAAAAAAAGAAAGAGAGGCGCGGGGTGGTGGCTACACCTGTATTC	1860
Dd		1801	AAAAAAAAAAAAAAAAAGAAAGAGAGGCGCGGGGTGGTGGCTACACCTGTATTC	1860
QY		1861	CAGACTTTTGGGAGGCGGAGCGCGGCGGATCACCTGTGTGACAGAGTTTGAACACAGCT	1920
Dd		1861	CAGACTTTTGGGAGGCGGAGCGCGGCGGATCACCTGTGTGACAGAGTTTGAACACAGCT	1920
QY		1921	GGCCACATGCGCAAAACCCCGCTGTACTCAAAATGCAAAAATTAAGCCAGGCGTGTAGC	1980
Dd		1921	GGCCACATGCGCAAAACCCCGCTGTACTCAAAATGCAAAAATTAAGCCAGGCGTGTAGC	1980
QY		1991	AGGACACCTATATCCCAGTACTTGGGAGGCTGGAGAGAGAAATGCTTGAACCCACGGA	2040
Dd		1991	AGGACACCTATATCCCAGTACTTGGGAGGCTGGAGAGAGAAATGCTTGAACCCACGGA	2040
QY		2041	GGTGAGAGTTSCATTAAGCTGAGATTCGTGCCGTGCACTCAGGCTTGGGAGACAAGGCA	2100
Dd		2041	GGTGAGAGTTSCATTAAGCTGAGATTCGTGCCGTGCACTCAGGCTTGGGAGACAAGGCA	2100
QY		2101	AGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAAATATTTG	2160
Dd		2101	AGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAAATATTTG	2160
QY		2161	GGAAGAGAGAGATGGGAGAGCATTTGCAGGAAATTTGTCTTATTCACAAAATGTAAAGCA	2220
Dd		2161	GGAAGAGAGAGATGGGAGAGCATTTGCAGGAAATTTGTCTTATTCACAAAATGTAAAGCA	2220

QY 2221 GCCAATAAGGATGCCATTTGTCCTTTGGTGTCTATTTGCCCAACAACTGCTTT 2280  
DB 2221 GCCAATAAGGATGCCATTTGTCCTTTGGTGTCTATTTGCCCAACAACTGCTTT 2280  
QY 2281 GACAGTGAAGAAATATTCAGAAATACCATATCCCTGCGTATTTACCTAGCAACCT 2340  
DB 2281 GACAGTGAAGAAATATTCAGAAATACCATATCCCTGCGTATTTACCTAGCAACCT 2340  
QY 2341 TGCATGAAGTGAAGATCCAGATCCAGAGAAACTGAAATGACAACTGCTATTTATC 2400  
DB 2341 TGCATGAAGTGAAGATCCAGATCCAGAGAAACTGAAATGACAACTGCTATTTATC 2400  
QY 2401 TTATTTGATCAATTTGTTGAAGAGTTAAATTTGTTACTTCATGATTTATTTAT 2460  
DB 2401 TTATTTGATCAATTTGTTGAAGAGTTAAATTTGTTACTTCATGATTTATTTAT 2460  
QY 2461 TTATTTATTTTGGCTGAATGATTTTATTAACATGATTTCTTTCTGATATATGA 2520  
DB 2461 TTATTTATTTTGGCTGAATGATTTTATTAACATGATTTCTTTCTGATATATGA 2520  
QY 2521 AATGAGTCTCAAGGCTTCATTAATTTATACCTTTAGAAATGATTTCAATACAGCTAT 2580  
DB 2521 AATGAGTCTCAAGGCTTCATTAATTTATACCTTTAGAAATGATTTCAATACAGCTAT 2580  
QY 2581 GTAATTTGATCAATTTGCTGATAGTGTCTACGAGCCATTTCTCTGATTTTGTAACT 2640  
DB 2581 GTAATTTGATCAATTTGCTGATAGTGTCTACGAGCCATTTCTCTGATTTTGTAACT 2640  
QY 2641 TTTATGACAGCAATTTGCTGCTGCTGCTCACTTCAATCACTTAATTAAGATTAAT 2700  
DB 2641 TTTATGACAGCAATTTGCTGCTGCTGCTGCTCACTTCAATCACTTAATTAAGATTAAT 2700  
QY 2701 TTTGGAAGCTGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760  
DB 2701 TTTGGAAGCTGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760  
QY 2761 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2818  
DB 2761 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2818

RESULT 2  
US-09-966-880a-10  
; Sequence 10, Application US/09966880A  
; Patient No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 6564  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880a-10

Query Match 77.1%; Score 2174; DB 9; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 GCCCTGATGAGTGTAGTACTAGAGAGCATTTGCTGACTTTGGAGCTTGATGCA 681

DB 3740 GCCCTGATGAGTGTAGTACTAGAGAGCATTTGCTGACTTTGGAGCTTGATGCA 3799  
QY 682 ACTTCAGGAATGTCAACACAGATGAATATCTGCTGAAGACAGTGAATAAAAACAG 741  
DB 3800 ACTTCAGGAATGTCAACACAGATGAATATCTGCTGAAGACAGTGAATAAAAACAG 3859  
QY 742 TCCCTCAAGTCTTCTGCTTTTATTTCTCAACTCTCACTTTCTTAGAGTTTACGAAA 801  
DB 3860 TCCCTCAAGTCTTCTGCTTTTATTTCTCAACTCTCACTTTCTTAGAGTTTACGAAA 3919  
QY 802 AATATTATATAGACTCTTTTAAAGATCTATGCTGAATATGAAGAAGAACAG 861  
DB 3920 AATATTATATAGACTCTTTTAAAGATCTATGCTGAATATGAAGAAGAACAG 3979  
QY 862 TCGGCCAGGAGAGTCTGCAATTTGCTGAGTTTGAATGCAACTTCTCCCTACTGG 921  
DB 3980 TCGGCCAGGAGAGTCTGCAATTTGCTGAGTTTGAATGCAACTTCTCCCTACTGG 4039  
QY 922 AATAACAGACTGCAGAGACTGGAGCATCTTAAGTGTCAACGTTTTCTATGACTTT 981  
DB 4040 AATAACAGACTGCAGAGACTGGAGCATCTTAAGTGTCAACGTTTTCTATGACTTT 4099  
QY 982 AGGTAGATGAAGAGCAAGAGTATCTTAAGACATGCTGAGAGATCAATGTTTT 1041  
DB 4100 AGGTAGATGAAGAGCAAGAGTATCTTAAGACATGCTGAGAGATCAATGTTTT 4159  
QY 1042 ATATCAACATCCCTTTTATTTGATTCATTTAGATTAACAGAGTGTAGATGATTT 1101  
DB 4160 ATATCAACATCCCTTTTATTTGATTCATTTAGATTAACAGAGTGTAGATGATTT 4219  
QY 1102 TTCTATCTTTTCCCTGAGCTTTACTTTCAAGTAAACAACTCTCCATCAGCCATG 1161  
DB 4220 TTCTATCTTTTCCCTGAGCTTTACTTTCAAGTAAACAACTCTCCATCAGCCATG 4279  
QY 1162 ATCTATAGAGCTCTTAATGAGATATCTGGGTGATTTGACCCCAACATCTCCAA 1221  
DB 4280 ATCTATAGAGCTCTTAATGAGATATCTGGGTGATTTGACCCCAACATCTCCAA 4339  
QY 1222 AGCATTAATTCATCATGCGCTGATGTTTAAATCAACAGATGTTTATGTTT 1281  
DB 4340 AGCATTAATTCATCATGCGCTGATGTTTAAATCAACAGATGTTTATGTTT 4399  
QY 1282 GTCAAAAGAGATTTGTATGAGTGGGATGAGATATAGACCATGATGATCTCA 1341  
DB 4400 GTCAAAAGAGATTTGTATGAGTGGGATGAGATATAGACCATGATGATCTCA 4459  
QY 1342 AGCTACTTAATTAAGATCTTTAAATGAGGAGAGAGACTGTGAACAGACCTTA 1401  
DB 4460 AGCTACTTAATTAAGATCTTTAAATGAGGAGAGAGACTGTGAACAGACCTTA 4519  
QY 1402 ATGGGTTGATGCTGATAGTGAATCTCTGGAACGCAACTCTTTAAGAGTCC 1461  
DB 4520 ATGGGTTGATGCTGATAGTGAATCTCTGGAACGCAACTCTTTAAGAGTCC 4579  
QY 1462 TAATTTAGAAACCCCAAACTTCATATCATATTTAGCAACATTTGGAAGAGTT 1521  
DB 4580 TAATTTAGAAACCCCAAACTTCATATCATATTTAGCAACATTTGGAAGAGTT 4639  
QY 1522 GCTTGAATGTTGGGAGAGAGAAATCTATGCTGCTGCTGCTCTTCAATCTCAGAA 1581  
DB 4640 GCTTGAATGTTGGGAGAGAGAAATCTATGCTGCTGCTGCTCTTCAATCTCAGAA 4699  
QY 1582 GCCAATCAGTCAAGTTTGTCTATATTTGTATGTTGTGATGCTTCCCAAGGTATA 1641  
DB 4700 GCCAATCAGTCAAGTTTGTCTATATTTGTATGTTGTGATGCTTCCCAAGGTATA 4759  
QY 1642 TTAACTATATTAAGAGTTGTGACAAAGATATTAAGTGGCAACGAGGCAAG 1701  
DB 4760 TTAACTATATTAAGAGTTGTGACAAAGATATTAAGTGGCAACGAGGCAAG 4819  
QY 1702 CTCATAGTCTAGCTCTTGGAGGTTGAGAGAGAGATGCTTGAACACAGGTGTTCA 1761

Db 4820 CTCATAGTCTAGCTGCTGGAGAGTGGAGAGGAGATGCTTGAACACAGCTGTCA 4879  
 QY 1762 AGGCCACCTGGGCAACATACAAAGATCCTGTCTCTAAAAAAGAA 1821  
 Db 4880 AGGCCACCTGGGCAACATACAAAGATCCTGTCTCTAAAAAAGAA 4939  
 QY 1822 AGAGAGAGGCGCGGCTGGTGGCTCAGCGCTGTAATCCAGACCTTGGAGCGGAGC 1881  
 Db 4940 AGAGAGAGGCGCGGCTGGTGGCTCAGCGCTGTAATCCAGACCTTGGAGCGGAGC 4999  
 QY 1882 CGGGCGGATCACTGTGTGAGAGAGTGGAGAGGCTGGAGAGTGGAGAGAGAGAG 1941  
 Db 5000 CGGGCGGATCACTGTGTGAGAGAGTGGAGAGGCTGGAGAGTGGAGAGAGAGAG 5059  
 QY 1942 TCTGTACTCAAAATGCAAAATTAAGCAGCGGTGTGAGAGAGAGTGTATCCAGCTA 2001  
 Db 5060 TCTGTACTCAAAATGCAAAATTAAGCAGCGGTGTGAGAGAGAGTGTATCCAGCTA 5119  
 QY 2002 CTGGGAGGCTGGAGAGAGAGTGTATCCAGAGAGTGTGGAGTGTAGAGTGTAGAGTGT 2061  
 Db 5120 CTGGGAGGCTGGAGAGAGAGTGTATCCAGAGAGTGTGGAGTGTAGAGTGTAGAGTGT 5179  
 QY 2062 AGATCGTGGCTGGAGAGAGTGTATCCAGAGAGTGTGGAGTGTAGAGTGTAGAGTGT 2121  
 Db 5180 AGATCGTGGCTGGAGAGAGTGTATCCAGAGAGTGTGGAGTGTAGAGTGTAGAGTGT 5239  
 QY 2122 AAAAAAG 2181  
 Db 5240 AAAAAAG 5299  
 QY 2182 TTGCAAG 2241  
 Db 5300 TTGCAAG 5359  
 QY 2242 GTCTCTTTGGTGTCTATTTGGCTTACAACTGTCTTACAGAGTGGAGAGAGAGAGAG 2301  
 Db 5360 GTCTCTTTGGTGTCTATTTGGCTTACAACTGTCTTACAGAGTGGAGAGAGAGAGAG 5419  
 QY 2302 AATAACATATCCCTGTGCGGTATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2361  
 Db 5420 AATAACATATCCCTGTGCGGTATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5479  
 QY 2362 CACGAG 2421  
 Db 5480 CACGAG 5539  
 QY 2422 AAGAGTAAAAATGTACTCATGTATCATTTATTTATTTATTTATTTATTTATTTATTT 2481  
 Db 5540 AAGAGTAAAAATGTACTCATGTATCATTTATTTATTTATTTATTTATTTATTTATTT 5599  
 QY 2482 GATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2541  
 Db 5600 GATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5659  
 QY 2542 AATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2601  
 Db 5660 AATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5719  
 QY 2602 TGGTGTACGAG 2661  
 Db 5720 TGGTGTACGAG 5779  
 QY 2662 CTGGCTCACTTCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2721  
 Db 5780 CTGGCTCACTTCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5839  
 QY 2722 ATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2781  
 Db 5840 ATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5899  
 QY 2782 TGGAAATAACTTGA 2795  
 Db 5900 TGGAAATAACTTGA 5913

RESULT 3  
 US-09-966-880A-35  
 ; Sequence 35, Application US/09966880A  
 ; Patent No. US20020164743A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Honjo, Tasuku  
 ; APPLICANT: Muramatsu, Masamichi  
 ; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
 ; FILE REFERENCE: 06501-088001  
 ; CURRENT APPLICATION NUMBER: US/09/966, 880A  
 ; CURRENT FILING DATE: 2001-09-28  
 ; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
 ; PRIOR FILING DATE: 2000-03-28  
 ; PRIOR APPLICATION NUMBER: JP 11-371382  
 ; PRIOR FILING DATE: 1999-12-27  
 ; PRIOR APPLICATION NUMBER: JP 11-178999  
 ; PRIOR FILING DATE: 1999-06-24  
 ; PRIOR APPLICATION NUMBER: JP 11-87192  
 ; PRIOR FILING DATE: 1999-03-29  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 35  
 ; LENGTH: 11204  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-966-880A-35

Query Match 77.1%; Score 2174; DB 9; Length 11204;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Db 622 GCCCTGTATGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 681  
 QY 8955 GCCCTGTATGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9014  
 Db 682 ACTTCAGAGATGTCACACAG 741  
 QY 9015 ACTTCAGAGATGTCACACAG 9074  
 Db 742 TCCCTCAAGTCTTCTCTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 801  
 QY 9075 TCCCTCAAGTCTTCTCTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 9134  
 Db 802 AATATTATATGAGTCTCTTTAAAGATCTATGCTTGAAGATGAGAGAGAGAGAGAGAG 861  
 QY 9135 AATATTATATGAGTCTCTTTAAAGATCTATGCTTGAAGATGAGAGAGAGAGAGAGAG 9194  
 Db 862 TCTGGCCAGGAGAGTGTGCAATTTGTCAGATTTGTAATGCAACATTTGCCCTACTGGG 921  
 QY 9195 TCTGGCCAGGAGAGTGTGCAATTTGTCAGATTTGTAATGCAACATTTGCCCTACTGGG 9254  
 Db 922 AATAACAGAGTGTGAG 981  
 QY 9255 AATAACAGAGTGTGAG 9314  
 Db 982 AGGTAG 1041  
 QY 9315 AGGTAG 9374  
 Db 1042 AATACACATCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1101  
 QY 9375 AATACACATCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 9434  
 Db 1102 TTTCTATTTCTTTCCCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 1161  
 QY 9435 TTTCTATTTCTTTCCCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 9494  
 Db 1162 ATCTATAGAGAGTCTATGAGAGATCTGGGTGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1221  
 QY 9495 ATCTATAGAGAGTCTATGAGAGATCTGGGTGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 9554

QY 1222 AGCATTAATCCATATGCGCTGTATGTTTATATCAGCAGAAAGATGTTTTANGTTT 1281  
DB 9555 AGCATTTATATCCATATGCGCTGTATGTTTATATCAGCAGAAAGATGTTTTANGTTT 9614  
QY 1282 GTACAAAAGAAAGATTGTATGGGTGGGATGGAGGTATAGACCATGATGGTACCTTCA 1341  
DB 9615 GTACAAAAGAAAGATTGTATGGGTGGGATGGAGGTATAGACCATGATGGTACCTTCA 9674  
QY 1342 AGCTACTTTAATAAGATCTTTAAATGGCAGAGAGACTGTGAACAGACCCCTATA 1401  
DB 9675 AGCTACTTTAATAAGATCTTTAAATGGCAGAGAGACTGTGAACAGACCCCTATA 9734  
QY 1402 ATGGGTGTAGTGTGAAGTGAACAATCTCTGGAACGCAACTCTTTAAGAAAGTCC 1461  
DB 9735 ATGGGTGTAGTGTGAAGTGAACAATCTCTGGAACGCAACTCTTTAAGAAAGTCC 9794  
QY 1462 TAATTTTGAACACCCCAAACTTCATATCATATTAAGCAAAATTTGGAAGAAATT 1521  
DB 9795 TAATTTTGAACACCCCAAACTTCATATCATATTAAGCAAAATTTGGAAGAAATT 9854  
QY 1522 GCTTGAATGTGGGAGAGAAATATATGGCTCTGTGGGTCTCTTCATCCAGAAAT 1581  
DB 9855 GCTTGAATGTGGGAGAGAAATATATGGCTCTGTGGGTCTCTTCATCCAGAAAT 9914  
QY 1582 GCCAATCAGGTCAAGGTTTGTCTACATTTGTATGTGTATGCTTCTCCCAAGATATA 1641  
DB 9915 GCCAATCAGGTCAAGGTTTGTCTACATTTGTATGTGTATGCTTCTCCCAAGATATA 9974  
QY 1642 TTAATCTATATAGAGAGTGTGTGACAAACAGATGATTAAGCTGCGAACCGTGACACG 1701  
DB 9975 TTAATCTATATAGAGAGTGTGTGACAAACAGATGATTAAGCTGCGAACCGTGACACG 10034  
QY 1702 CTCATAGTGTAGCTGTGGGAGGTTGAGAGGAGAGAGTGGCTTGACACAGTGTCA 1761  
DB 10035 CTCATAGTGTAGCTGTGGGAGGTTGAGAGGAGAGAGTGGCTTGACACAGTGTCA 10094  
QY 1762 AGGCCAGCTGGGCAACATACAAAGATCTGTCTCTCAAAAAAAGAA 1821  
DB 10095 AGGCCAGCTGGGCAACATACAAAGATCTGTCTCTCAAAAAAAGAA 10154  
QY 1822 AGAGAGAGGGCCGGCGTGTGTGCTCAAGCCTGTATCCAGCACTTTGGGAGCCGAGC 1881  
DB 10155 AGAGAGAGGGCCGGCGTGTGTGCTCAAGCCTGTATCCAGCACTTTGGGAGCCGAGC 10214  
QY 1882 CGGGCGGATACCTGTGTGTCAGAGGTTTGAGACCAAGCTGCGCAACATGCGAAACCCCG 1941  
DB 10215 CGGGCGGATACCTGTGTGTCAGAGGTTTGAGACCAAGCTGCGCAACATGCGAAACCCCG 10274  
QY 1942 TCTGTACTCAAAATGCAAAATTTAGCAGGCGTGTGAGCAGGCACTGTATCCAGCTA 2001  
DB 10275 TCTGTACTCAAAATGCAAAATTTAGCAGGCGTGTGAGCAGGCACTGTATCCAGCTA 10334  
QY 2002 CTTGGAGGCTGAGCAGAGAAATCGCTTAACCCAGAGAGTGGAGTTTGCATTAAGCTG 2061  
DB 10335 CTTGGAGGCTGAGCAGAGAAATCGCTTAACCCAGAGAGTGGAGTTTGCATTAAGCTG 10394  
QY 2062 AGATCGTGGCGTGTGCTCAAGCTGTGGGCAACAAGAGTGTGTCTCAGAAAAAAA 2121  
DB 10395 AGATCGTGGCGTGTGCTCAAGCTGTGGGCAACAAGAGTGTGTCTCAGAAAAAAA 10454  
QY 2122 AAAAAAAG 2181  
DB 10455 AAAAAAAG 10514  
QY 2182 TTGCAAGAGAAATTTGCTTTATCCAAACAATGTAAAGAGCAATAGGATCCCTATT 2241  
DB 10515 TTGCAAGAGAAATTTGCTTTATCCAAACAATGTAAAGAGCAATAGGATCCCTATT 10574  
QY 2242 GTCTCTTTTGTGTCTATTTGTCCCTAACAACTGTCTTTGACAGTGAAGAAAAATTTCA 2301  
DB 10575 GTCTCTTTTGTGTCTATTTGTCCCTAACAACTGTCTTTGACAGTGAAGAAAAATTTCA 10634  
QY 2302 AATAACCAATCCCTGTGGCGTTATTTACTAGCAACCCCTTGCAATGAAGATGAGCAGATC 2361

DB 10635 AATAACCAATCCCTGTGGCGTTATTTACTTACCAACCCCTGCAATGAAGATGAGCAGATC 10694  
QY 2362 CACAGGAAACCTTGAATGACACAGCTGCTTATTTAACTTTGTACATAAGTTGTA 2421  
DB 10695 CACAGGAAACCTTGAATGACACAGCTGCTTATTTAACTTTGTACATAAGTTGTA 10754  
QY 2422 AAGAGTAAACCTTGAATGACACAGCTGCTTATTTAACTTTGTACATAAGTTGTA 2481  
DB 10755 AAGAGTAAACCTTGAATGACACAGCTGCTTATTTAACTTTGTACATAAGTTGTA 10814  
QY 2482 GATTTTATTAACATGATTTCTTTTCTGATATATTTGAATGAGTCTCAAGCTTCAT 2541  
DB 10815 GATTTTATTAACATGATTTCTTTTCTGATATATTTGAATGAGTCTCAAGCTTCAT 10874  
QY 2542 AATTTATTAACATGATTTCTTTTCTGATATATTTGAATGAGTCTCAAGCTTCAT 2601  
DB 10875 AATTTATTAACATGATTTCTTTTCTGATATATTTGAATGAGTCTCAAGCTTCAT 10934  
QY 2602 TGGTGTACGAGCCATTTCTGTGATTTTGTAACTTTATGACAGCAATTTGCTT 2661  
DB 10935 TGGTGTACGAGCCATTTCTGTGATTTTGTAACTTTATGACAGCAATTTGCTT 10994  
QY 2662 CTGGCTCACTTCAATCAGTTAAATTAATGATTAATTTTGAAGCTGTGAAGATATA 2721  
DB 10995 CTGGCTCACTTCAATCAGTTAAATTAATGATTAATTTTGAAGCTGTGAAGATATA 11054  
QY 2722 ATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2781  
DB 11055 ATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11114  
QY 2782 TGGATTAATCTGA 2795  
DB 11115 TGGATTAATCTGA 11128

RESULT 4  
US-09-966-880A-15  
; Sequence 15, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1998-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; LENGTH: 2172  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880A-15

Query Match 77.1%; Score 2172; DB 9; Length 2172;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 CCCCTGTATGAGTGTGATGACTTACGAGAGCATTCTTCTACTTTGGACTTTGATAGCA 682  
DB 1 CCCCTGTATGAGTGTGATGACTTACGAGAGCATTCTTCTACTTTGGACTTTGATAGCA 60  
QY 683 CTTCAAGGATGTCAACAGATGAATATCTCTGTAAGAGAGTGAATTAACAGT 742

Dh 61 CTCCAGGATGTCAACACAGATGAATATCTGCTGAAGACAGTGGATAAAAACGT 120  
Qy 743 CCTTCAGTCTTCTCTGTTTTATTTCTTCAACTCTCATTTCTTAGGTTTACAGAAAA 802  
Db 121 CCTTCAGTCTTCTCTGTTTTATTTCTTCAACTCTCATTTCTTAGGTTTACAGAAAA 180  
Qy 803 ATATTATATGAGCTCTTTAAAAAGATCTATGCTGAAATATGAAAGAAACAGAGT 862  
Db 181 ATATTATATGAGCTCTTTAAAAAGATCTATGCTGAAATATGAAAGAAACAGAGT 240  
Qy 863 CTGGCCAGGAGCGTCTGCAATTGCTGCAAGTTTGAATGCAACATTTGCCCTACTGGGA 922  
Db 241 CTGGCCAGGAGCGTCTGCAATTGCTGCAAGTTTGAATGCAACATTTGCCCTACTGGGA 300  
Qy 923 ATACAGAACTGCGAGGACCTGGGAGCATCTCTAAAGTGTCAACGTTTTCTATGACTTTA 982  
Db 301 ATACAGAACTGCGAGGACCTGGGAGCATCTCTAAAGTGTCAACGTTTTCTATGACTTTA 360  
Qy 983 GGTAGATGAGAGAGAGATGATCTCTAAAAAGCATGTGAGAGATCAAAATGTTTTA 1042  
Db 361 GGTAGATGAGAGAGAGATGATCTCTAAAAAGCATGTGAGAGATCAAAATGTTTTA 420  
Qy 1043 TATCAACATCCTTTATTTATTTGATTCATTTGATTTAAGTTAGTTAGATGATTTT 1102  
Db 421 TATCAACATCCTTTATTTATTTGATTCATTTGATTTAAGTTAGTTAGATGATTTT 480  
Qy 1103 TCTATTTCTTTCCCTTGACGTTTACTTCAAGTATACAAACCTCTTCATCAGGCCATGA 1162  
Db 481 TCTATTTCTTTCCCTTGACGTTTACTTCAAGTATACAAACCTCTTCATCAGGCCATGA 540  
Qy 1163 TCTATGAGACCTCTTATGAGATGATCTGGGTGATTTGACCCCAACCATCTCTCCAA 1222  
Db 541 TCTATGAGACCTCTTATGAGATGATCTGGGTGATTTGACCCCAACCATCTCTCCAA 600  
Qy 1223 GCATTAATATCCAACTCATGCGCTGATTTTAACTGAGAGAGCATGTTTTATGTTG 1282  
Db 601 GCATTAATATCCAACTCATGCGCTGATTTTAACTGAGAGAGCATGTTTTATGTTG 660  
Qy 1283 TACAAAAGAGATTTGATGAGGTGGGATGAGGATATGACCATGATGTCACCTCAA 1342  
Db 661 TACAAAAGAGATTTGATGAGGTGGGATGAGGATATGACCATGATGTCACCTCAA 720  
Qy 1343 GCTACTTAATAAGATCTTAAATGGGAGAGAGCATGTGAACAAGAACCTTATTA 1402  
Db 721 GCTACTTAATAAGATCTTAAATGGGAGAGAGCATGTGAACAAGAACCTTATTA 780  
Qy 1403 TGGGTGATGCTGAAGTAGCAAACTCTTGGAAGCAACCTTTTAAAGAGTCCCT 1462  
Db 781 TGGGTGATGCTGAAGTAGCAAACTCTTGGAAGCAACCTTTTAAAGAGTCCCT 840  
Qy 1463 AATTAGAAACCCCAAACTTCATATCATTAATTAGCAAACAATTGGAGAGAAATG 1522  
Db 841 AATTAGAAACCCCAAACTTCATATCATTAATTAGCAAACAATTGGAGAGAAATG 900  
Qy 1523 CTGAATGTTGGGAGAGGAAATCTATGGCTCTGCTGCTCTCATCTCAGAAATG 1582  
Db 901 CTGAATGTTGGGAGAGGAAATCTATGGCTCTGCTGCTCTCATCTCAGAAATG 960  
Qy 1583 CCAATCAGTCAAGGTTGCTACATTTGATGCTGATGCTGCTGCCAAAGTATAT 1642  
Db 961 CCAATCAGTCAAGGTTGCTACATTTGATGCTGATGCTGCTGCCAAAGTATAT 1020  
Qy 1643 TAACTATATAGAGAGTGTGACAAAACAGATATGATTAAGCTCGAAGCTGGCACAGC 1702  
Db 1021 TAACTATATAGAGAGTGTGACAAAACAGATATGATTAAGCTCGAAGCTGGCACAGC 1080  
Qy 1703 TCAATGTTTACGCTGCTTGGAGGTTGAGAGGGAGGATGGCTTGAACACAGGTGTTCAA 1762  
Db 1081 TCAATGTTTACGCTGCTTGGAGGTTGAGAGGGAGGATGGCTTGAACACAGGTGTTCAA 1140  
Qy 1763 GGCAGAGCTGGGCAACATTAACAGATCTCTCTCAAAAAAAGAAAAAAGAAAA 1822  
Db 1141 GGCAGAGCTGGGCAACATTAACAGATCTCTCTCAAAAAAAGAAAAAAGAAAA 1200

Qy 1823 GAGAGAGGCGCGGCGTGTGCTCAAGCTGTATATCCAGACCTTTGGAGCCGAGCC 1882  
Db 1201 GAGAGAGGCGCGGCGTGTGCTCAAGCTGTATATCCAGACCTTTGGAGCCGAGCC 1260  
Qy 1883 GGGCGGATACCTGAGTGTGAGAGTTTGAAGACACCTGGCCAAACATGGCAAAACCCCGT 1942  
Db 1261 GGGCGGATACCTGAGTGTGAGAGTTTGAAGACACCTGGCCAAACATGGCAAAACCCCGT 1320  
Qy 1943 CTGTACTCAAAATGCAAAATATGACAGCGGTGTGAGAGGACCTGTAAATCCAGCTAC 2002  
Db 1321 CTGTACTCAAAATGCAAAATATGACAGCGGTGTGAGAGGACCTGTAAATCCAGCTAC 1380  
Qy 2003 TTGGAGGCTGAGGACAGAGAAATGCTTGAACCCAGAGGTGAGAGTTGCAATGCTGA 2062  
Db 1381 TTGGAGGCTGAGGACAGAGAAATGCTTGAACCCAGAGGTGAGAGTTGCAATGCTGA 1440  
Qy 2063 GATCGTCCGCTGCACTCCAGCCTGGGAGCAAGAGCAAGACTGCTCAGAAAAAAA 2122  
Db 1441 GATCGTCCGCTGCACTCCAGCCTGGGAGCAAGAGCAAGACTGCTCAGAAAAAAA 1500  
Qy 2123 AAAAAAAGAT 2182  
Db 1501 AAAAAAAGAT 1560  
Qy 2183 TGCAGAGAAATGTGCTTTATCCAACAAAATGTAAAGAGCCAAATAGGAGATCCCTATTG 2242  
Db 1561 TGCAGAGAAATGTGCTTTATCCAACAAAATGTAAAGAGCCAAATAGGAGATCCCTATTG 1620  
Qy 2243 TCTCTTTGCTGCTATTTGCTCCCTACACACGCTGTTTGAGAGAGAGAGAGAGAGAG 2302  
Db 1621 TCTCTTTGCTGCTATTTGCTCCCTACACACGCTGTTTGAGAGAGAGAGAGAGAGAG 1680  
Qy 2303 ATACCATATCCCTGCTGCGCTATTAACCTAGAGACCTTGAATTAATGAGAGAGATCC 2362  
Db 1681 ATACCATATCCCTGCTGCGCTATTAACCTAGAGACCTTGAATTAATGAGAGAGATCC 1740  
Qy 2363 ACAGGAAACCTGAAATGCAACACTGCTATTTAATCTTAATGATGACATAAGTTGTAA 2422  
Db 1741 ACAGGAAACCTGAAATGCAACACTGCTATTTAATCTTAATGATGACATAAGTTGTAA 1800  
Qy 2423 AGAGTTAAAAATGTTACTCTCATGATTCATTTATTTATTTATTTTCCGCTATG 2482  
Db 1801 AGAGTTAAAAATGTTACTCTCATGATTCATTTATTTATTTATTTTCCGCTATG 1860  
Qy 2483 ATTTTATTAACATGATTTCCCTTCTGATATTAATGAAAGAGCTCAAGCTCATA 2542  
Db 1861 ATTTTATTAACATGATTTCCCTTCTGATATTAATGAAAGAGCTCAAGCTCATA 1920  
Qy 2543 AATTATTAACCTTAAGATGATTTCTAATTAACAAGCTATGTAATGTAACATGAGTAT 2602  
Db 1921 AATTATTAACCTTAAGATGATTTCTAATTAACAAGCTATGTAATGTAACATGAGTAT 1980  
Qy 2603 GGTGCTAGCAACCATTTCTGATTTTATTAACCTTTATGACACCAATTTGCTTC 2662  
Db 1981 GGTGCTAGCAACCATTTCTGATTTTATTAACCTTTATGACACCAATTTGCTTC 2040  
Qy 2663 TGGCTACCTTCAACAGTTAATATGATTAATTAATTTTGAAGCTGTGAAGATAAA 2722  
Db 2041 TGGCTACCTTCAACAGTTAATATGATTAATTAATTTTGAAGCTGTGAAGATAAA 2100  
Qy 2723 TACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2782  
Db 2101 TACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
Qy 2783 GGAATTAACCTG 2794  
Db 2161 GGAATTAACCTG 2172

RESULT 5  
US-09-966-880A-13  
; Sequence 13, Application US/09966880A

```
Patent No. US2002016473A1
GENERAL INFORMATION:
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966,880A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: JP 11-178999
PRIOR FILING DATE: 1999-06-24
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 271
TYPE: DNA
ORGANISM: Homo sapiens
US-09-966-880A-13

Query Match
Best Local Similarity 100.0%; Score 271; DB 9; Length 271;
Pred. No. 2,4e-126;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 AACGGCTGCCACGCTGGAAATTCCTCTCCGCTACATCTCGACCTGGGACCTAGACCT 295
DB 1 AACGGCTGCCACGCTGGAAATTCCTCTCCGCTACATCTCGACCTGGGACCTAGACCT 60
QY 296 GGGCGCTGCTACCGCGTACCTGGTACCTCCCTGGAGCCCTCTCGACTGTGCCGA 355
DB 61 GGGCGCTGCTACCGCGTACCTGGTACCTCCCTGGAGCCCTCTCGACTGTGCCGA 120
QY 356 CATGTGGCCGACTTCTGCGAGGAGACCCCACTCAGTCTGAGGATCTTCACCGCGC 415
DB 121 CATGTGGCCGACTTCTGCGAGGAGACCCCACTCAGTCTGAGGATCTTCACCGCGC 180
QY 416 CTCTACTCTGTGAGGAGCCGACGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGGCC 475
DB 181 CTCTACTCTGTGAGGAGCCGACGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGGCC 240
QY 476 GGGGTGCAATATGCAATCATGACCTTCAAG 506
DB 241 GGGGTGCAATATGCAATCATGACCTTCAAG 271

RESULT 6
US-09-966-880A-12
Sequence 12, Application US/09966880A
Patent No. US2002016473A1
GENERAL INFORMATION:
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966,880A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: JP 11-178999
PRIOR FILING DATE: 1999-06-24
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 148
TYPE: DNA
```

```
ORGANISM: Homo sapiens
US-09-966-880A-12
Query Match
Best Local Similarity 100.0%; Score 148; DB 9; Length 148;
Pred. No. 3.3e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 CCTCTTGATGAACCGGAGAGTTCTTACCAATTCAAAATGTCGCTGGGCTTAAGG 147
DB 1 CCTCTTGATGAACCGGAGAGTTCTTACCAATTCAAAATGTCGCTGGGCTTAAGG 60
QY 148 TCGGCGTGAACCTACCTTGTCTAGCTAGTAAGAGGCTGACAGTCTTCTTTC 207
DB 61 TCGGCGTGAACCTACCTTGTCTAGCTAGTAAGAGGCTGACAGTCTTCTTTC 120
QY 208 ACTGACTTTGTTATCTTCCGCAATAG 235
DB 121 ACTGACTTTGTTATCTTCCGCAATAG 148

RESULT 7
US-09-966-880A-11
Sequence 11, Application US/09966880A
Patent No. US2002016473A1
GENERAL INFORMATION:
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966,880A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: JP 11-178999
PRIOR FILING DATE: 1999-06-24
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 87
TYPE: DNA
ORGANISM: Homo sapiens
US-09-966-880A-11

Query Match
Best Local Similarity 100.0%; Score 87; DB 9; Length 87;
Pred. No. 2.1e-33;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGACCATCATTAATTAAGTGAATTTTCTGGCCCTGAGACTGCGAGGAGGCAAGA 60
DB 1 AGAAGACCATCATTAATTAAGTGAATTTTCTGGCCCTGAGACTGCGAGGAGGCAAGA 60
QY 61 AGAAGACTGTGAGACACCATATGACAG 87
DB 61 AGAAGACTGTGAGACACCATATGACAG 87

RESULT 8
US-09-966-880A-9
Sequence 9, Application US/09966880A
Patent No. US2002016473A1
GENERAL INFORMATION:
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966,880A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
```

PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 5514  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (1)...(1031)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (1032)...(1118)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (1119)...(5514)  
US-09-966-880A-9

Query Match 3.1%; Score 87; DB 9; Length 5514;  
Best Local Similarity 100.0%; Pred. No. 3.2e-33;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGAAGACCATCATTAATGAGAGAGATTTTCTGGCCCTGAGACTGCGAGGAGGAGAGA 60  
DB 1032 AGAAGACCATCATTAATGAGAGAGATTTTCTGGCCCTGAGACTGCGAGGAGGAGAGA 1091

OY 61 AGACACTGTGACACCACTATGACAG 87  
DB 1092 AGACACTGTGACACCACTATGACAG 1118

RESULT 9  
US-09-966-880A-14  
Sequence 14, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 116  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-14

Query Match 2.8%; Score 78; DB 9; Length 116;  
Best Local Similarity 100.0%; Pred. No. 7.6e-29;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 545 GAAAGAACTTCAAGGCTGGAGAGGCTGCATGAAATTCAGTCTGCTCCAGACAG 604  
DB 39 GAAAGAACTTCAAGGCTGGAGAGGCTGCATGAAATTCAGTCTGCTCCAGACAG 98  
OY 605 CTTGGGGCATCCTTTTG 622

DB 99 CTTGGGGCATCCTTTTG 116

RESULT 10  
US-09-962-436-281/c  
Sequence 281, Application US/09962436  
Patent No. US20020081301A1  
GENERAL INFORMATION:  
APPLICANT: Soppel, Daniel  
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign  
FILE REFERENCE: 689290-75  
CURRENT APPLICATION NUMBER: US/09/962, 436  
PRIOR FILING DATE: 2001-09-25  
PRIOR APPLICATION NUMBER: US/60/235, 082  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/234, 924  
PRIOR FILING DATE: 2000-09-25  
NUMBER OF SEQ ID NOS: 568  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 281  
LENGTH: 167343  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-962-436-281

Query Match 2.5%; Score 70; DB 10; Length 167343;  
Best Local Similarity 100.0%; Pred. No. 1.8e-24;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1986 CCTGTATCCCACTACTTGGAGGCTGAGCAGAGAGATGCTTGAACCCAGAGGTGG 2045  
DB 131153 CCTGTATCCCACTACTTGGAGGCTGAGCAGAGAGATGCTTGAACCCAGAGGTGG 131094

OY 2046 AGTTGCAGT 2055  
DB 131093 AGTTGCAGT 131084

RESULT 11  
US-09-964-824A-273/c  
Sequence 273, Application US/09964824A  
Patent No. US20020102531A1  
GENERAL INFORMATION:  
APPLICANT: Horrigan, Stephen  
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign  
FILE REFERENCE: 689290-73  
CURRENT APPLICATION NUMBER: US/09/964, 824A  
PRIOR FILING DATE: 2001-09-27  
PRIOR APPLICATION NUMBER: US/60/236, 033  
PRIOR FILING DATE: 2000-09-28  
PRIOR APPLICATION NUMBER: US/60/236, 032  
PRIOR FILING DATE: 2000-09-28  
PRIOR APPLICATION NUMBER: US/60/236, 028  
PRIOR FILING DATE: 2000-09-28  
NUMBER OF SEQ ID NOS: 583  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 273  
LENGTH: 167343  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-964-824A-273

Query Match 2.5%; Score 70; DB 10; Length 167343;  
Best Local Similarity 100.0%; Pred. No. 1.8e-24;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1986 CCTGTATCCCACTACTTGGAGGCTGAGCAGAGAGATGCTTGAACCCAGAGGTGG 2045  
DB 131153 CCTGTATCCCACTACTTGGAGGCTGAGCAGAGAGATGCTTGAACCCAGAGGTGG 131094  
OY 2046 AGTTGCAGT 2055



Db 131093 AGGTGCAGT 131084

RESULT 12  
US-10-074-095-922

Sequence 922, Application US/10074095  
Publication No. US2003007704A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC008C1

CURRENT APPLICATION NUMBER: US/10/074,095

PRIOR FILING DATE: 2002-02-14

PRIOR APPLICATION NUMBER: 09/764,860

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION NUMBER: 60/179,065

PRIOR FILING DATE: 2000-01-31

PRIOR APPLICATION NUMBER: 60/180,628

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: 60/214,886

PRIOR FILING DATE: 2000-06-28

PRIOR APPLICATION NUMBER: 60/217,487

PRIOR FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/225,758

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/220,963

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/217,496

PRIOR FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/225,447

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/218,290

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: 60/226,868

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: 60/216,647

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: 60/225,267

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/216,880

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: 60/225,270

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/251,869

PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/235,834

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: 60/234,274

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: 60/234,223

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: 60/228,924

PRIOR FILING DATE: 2000-08-30

PRIOR APPLICATION NUMBER: 60/224,518

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/236,369

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/224,519

PRIOR FILING DATE: 2000-08-14

PRIOR FILING DATE: 2000-11-01

PRIOR APPLICATION NUMBER: 60/225,268

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/236,368

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/251,856

PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/251,868

PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/229,344

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/234,997

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: 60/229,343

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/229,345

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/229,287

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/229,513

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: 60/231,413

PRIOR FILING DATE: 2000-09-08

PRIOR APPLICATION NUMBER: 60/229,509

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: 60/236,367

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/237,039

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/237,038

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/236,370

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/237,040

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/240,960

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/239,935

PRIOR FILING DATE: 2000-10-13

PRIOR APPLICATION NUMBER: 60/239,937

PRIOR FILING DATE: 2000-10-13

PRIOR APPLICATION NUMBER: 60/241,787

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/246,474

PRIOR FILING DATE: 2000-11-08

PRIOR APPLICATION NUMBER: 60/246,532

PRIOR FILING DATE: 2000-11-08

PRIOR APPLICATION NUMBER: 60/249,216

PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: 60/249,210

PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: 60/226,681

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: 60/225,759

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/225,213

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/227,182

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: 60/225,214

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/235,836

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: 60/230,438

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/215,135

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: 60/225,266

PRIOR FILING DATE: 2000-08-14

;; PRIOR APPLICATION NUMBER: 60/249,218  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,208  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,213  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,212  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,207  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,245  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,244  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,217  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,211  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,215  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,264  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,214  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,297  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/232,400  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/231,242  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,081  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,080  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,414  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,244  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/233,064  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/233,063  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,397  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,399  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,401  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/241,808  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,826  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,786  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,221  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,475  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/231,243  
;; PRIOR FILING DATE: 2000-09-08

Query Match 2.3%; Score 66; DB 9; Length 125;  
Best Local Similarity 100.0%; Pred. No. 8.8e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1990 TAAATCCAGCTACTTGGAGGCTGAGCGAGAGATGCTTGAACCCAGAGGTGAGGT 2049  
DB 1 TAAATCCAGCTACTTGGAGGCTGAGCGAGAGATGCTTGAACCCAGAGGTGAGGT 60  
QY 2050 TGCAGT 2055  
DB 61 TGCAGT 66

RESULT 13  
US-10-074-095-923  
;; Sequence 923, Application US/10074095  
;; Publication No. US20030077704A1  
;; GENERAL INFORMATION:  
;; Applicant: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
;; FILE REFERENCE: PC008C1  
;; CURRENT APPLICATION NUMBER: US/10/074,095  
;; CURRENT FILING DATE: 2002-02-14  
;; PRIOR APPLICATION NUMBER: 09/764,860  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 60/179,065  
;; PRIOR FILING DATE: 2000-01-31  
;; PRIOR APPLICATION NUMBER: 60/180,628  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: 60/214,886  
;; PRIOR FILING DATE: 2000-06-28  
;; PRIOR APPLICATION NUMBER: 60/217,487  
;; PRIOR FILING DATE: 2000-07-11  
;; PRIOR APPLICATION NUMBER: 60/225,758  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/220,963  
;; PRIOR FILING DATE: 2000-07-26  
;; PRIOR APPLICATION NUMBER: 60/217,496  
;; PRIOR FILING DATE: 2000-07-11  
;; PRIOR APPLICATION NUMBER: 60/225,447  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/218,290  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: 60/225,757  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/226,868  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/216,647  
;; PRIOR FILING DATE: 2000-07-07  
;; PRIOR APPLICATION NUMBER: 60/225,267  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/216,880  
;; PRIOR FILING DATE: 2000-07-07  
;; PRIOR APPLICATION NUMBER: 60/225,270  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/251,869  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/235,834  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/234,274  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: 60/234,223  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: 60/228,924  
;; PRIOR FILING DATE: 2000-08-30  
;; PRIOR APPLICATION NUMBER: 60/224,518  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/236,369  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/224,519  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/220,964  
;; PRIOR FILING DATE: 2000-07-26  
;; PRIOR APPLICATION NUMBER: 60/241,809  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/249,299  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/236,327  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/241,785  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/244,617  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/225,268  
;; PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208

PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232,400  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/231,242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,080  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,414  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,244  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,064  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/233,063  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,397  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,399  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,401  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/241,808  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,826  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,786  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,221  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,475  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/231,243  
PRIOR FILING DATE: 2000-09-08

Query Match 2.38; Score 66; DB 9; Length 125;  
Best Local Similarity 100.0%; Pred. No. 8.8e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1990 TAATCCAGCTACTGTGGAGGCTGAGGCGAGAGATCGCTTGACCCAGAGAGTGGAGGT 2049  
|||||  
Db 1 TAATCCAGCTACTGTGGAGGCTGAGGCGAGAGATCGCTTGACCCAGAGAGTGGAGGT 60  
Qy 2050 TGCAGT 2055  
|||||  
Db 61 TGCAGT 66

RESULT 14  
US-09-764-860-922

; Sequence 922, Application US/09764860  
; Patent No. US20020094953A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC008  
; CURRENT APPLICATION NUMBER: US/09/764,860  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1198  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 922  
; LENGTH: 125  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-860-922

Query Match 2.3%; Score 66; DB 10; Length 125;  
Best Local Similarity 100.0%; Pred. No. 8.8e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1990 TAATCCAGCTACTTGGAGGCTGAGGACAGAGATCGCTTGAACCCAGAGGTTGAGGT 2049  
|||||  
Db 1 TAATCCAGCTACTTGGAGGCTGAGGACAGAGATCGCTTGAACCCAGAGGTTGAGGT 60

QY 2050 TGCAGT 2055  
|||||  
Db 61 TGCAGT 66

RESULT 15  
US-09-764-860-923  
; Sequence 923, Application US/09764860  
; Patent No. US20020094953A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC008  
; CURRENT APPLICATION NUMBER: US/09/764,860  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1198  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 923  
; LENGTH: 125  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-860-923

Query Match 2.3%; Score 66; DB 10; Length 125;  
Best Local Similarity 100.0%; Pred. No. 8.8e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1990 TAATCCAGCTACTTGGAGGCTGAGGACAGAGATCGCTTGAACCCAGAGGTTGAGGT 2049  
|||||  
Db 1 TAATCCAGCTACTTGGAGGCTGAGGACAGAGATCGCTTGAACCCAGAGGTTGAGGT 60

QY 2050 TGCAGT 2055  
|||||  
Db 61 TGCAGT 66

Search completed: July 7, 2003, 23:28:03  
Job time : 493 secs





PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PERKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB: ABL05153.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 9942; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
SQ Sequence 1267 AA;  
  
Query Match 4.5%; Score 9; DB 22; Length 1267;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 117 EDRKAEPFG 125  
|||  
Db 237 EDRKAEPFG 245  
  
RESULT 4  
AAM30763  
ID AAM30763 standard; Protein: 1178 AA.  
XX  
AC AAM30763;  
XX  
DT 07-MAY-1998 (first entry)  
XX  
DE Mannose-1-phosphate transferase protein MNNA.  
XX  
KW Mannose-1-phosphate transfer; MNNA gene; enzyme; yeast; regulatory gene;  
KW human; high mannose type neutral saccharide chain.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN JPO9266792-A.  
PN  
PD 14-OCT-1997.  
XX  
PF 29-MAR-1996; 96JP-0075667.  
XX  
PR 29-MAR-1996; 96JP-0075667.  
XX  
PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
XX  
DR WPI: 1997-553460/51.  
DR N-PSDB: AAT91902.  
XX  
PT Positive regulatory gene of mannose-1-phosphate transfer in yeast -  
PT useful for high mannose type neutral saccharide chain production

XX  
PS Claim 1; Page 14-17; 23pp; Japanese.  
XX  
CC This sequence is encoded by the gene of the invention, designated MNNA,  
CC and is a protein which positively regulates mannose-1-phosphate  
CC transfer in yeast. The gene is useful for the preparation of human high  
CC mannose type neutral saccharide chain.  
XX  
SQ Sequence 1178 AA;  
  
Query Match 4.0%; Score 8; DB 18; Length 1178;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 176 LRRILPL 183  
|||  
Db 25 LRRILPL 32  
  
RESULT 5  
AAU50497  
ID AAU50497 standard; Protein: 50 AA.  
XX  
AC AAU50497;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #11393.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-20841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
PI  
DR WPI: 2001-616774/71.  
DR N-PSDB: AAS59548.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID NO 11692; 1069pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 50 AA;

SO Query Match 3.5%; Score 7; DB 22; Length 50;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 DAFRTLG 197  
DB 18 DAFRTLG 24

RESULT 6  
ABB32258  
ID ABB32258 standard; Peptide; 54 AA.

XX ABB32258;  
AC  
DT 01-FEB-2002 (first entry)

DE Peptide #4909 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer.

XX Homo sapiens.

OS  
XX  
PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000662.

PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-496933/54.

PT New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -

PS Claim 27; SEQ ID NO 15226; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for

CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 54 AA;

SO Query Match 3.5%; Score 7; DB 22; Length 54;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 138 IMTFKDY 144  
DB 6 IMTFKDY 12

RESULT 7  
ABB37512  
ID ABB37512 standard; Peptide; 54 AA.

XX ABB37512;  
AC  
DT 04-FEB-2002 (first entry)

DE Peptide #5018 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

OS  
XX  
PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000669.

PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -

PS Claim 27; SEQ ID NO 30147; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 54 AA;

Query Match 3.5%; Score 7; DB 22; Length 54;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 138 IMTFKDY 144  
 Db 6 IMTFKDY 12

RESULT 8  
 ABB22810  
 ID ABB22810 standard; Protein; 54 AA.  
 XX  
 AC ABB22810;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #4809 encoded by probe for measuring heart cell gene expression.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PE 30-JAN-2001; 2001WO-US00666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PS Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 PI WPI; 2001-488899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 XX  
 PS Claim 15; SEQ ID No 24580; 530pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 54 AA;

Query Match 3.5%; Score 7; DB 22; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 IMTFKDY 144  
 Db 6 IMTFKDY 12

RESULT 9  
 AAM58174  
 ID AAM58174 standard; Protein; 54 AA.

XX  
 AC AAM58174;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30279.  
 XX  
 KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PE 30-JAN-2001; 2001WO-US00667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PS Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 PI WPI; 2001-483446/52.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 XX  
 PS Example 4; SEQ ID NO: 30279; 650pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention.  
 XX  
 SQ Sequence 54 AA;

Query Match 3.5%; Score 7; DB 22; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 IMTFKDY 144  
 Db 6 IMTFKDY 12

RESULT 10  
 AAM70625  
 ID AAM70625 standard; Protein; 54 AA.  
 XX  
 AC AAM70625;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30931.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukemia; lymphoma; myeloma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157276-A2.

```

XX 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 30931; 658bp + Sequence Listing: English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX Sequence 54 AA;
XX
XX Query Match 3.5%; Score 7; DB 22; Length 54;
XX Best Local Similarity 100.0%; Pred. No. 23;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 IMTFKDY 144
DB 6 IMTFKDY 12

```

RESULT 11  
AAM18471  
ID AAM18471 standard; Protein: 54 AA.  
XX  
XX AAM18471;  
XX  
XX 12-OCT-2001 (first entry)  
XX  
XX Peptide #4905 encoded by probe for measuring cervical gene expression.  
XX Probe: human; microarray; gene expression; cervical epithelial cell;  
XX cervical cancer.  
XX  
XX Homo sapiens.  
XX  
XX WO200157278-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00670.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 27; SEQ ID NO 23297; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SEN; see A110068-A118459). The present sequence is a peptide encoded
XX by one such probe. The SENs are derived from human HeLa cells. The SENs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 54 AA;
XX
XX Query Match 3.5%; Score 7; DB 22; Length 54;
XX Best Local Similarity 100.0%; Pred. No. 23;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 IMTFKDY 144
DB 6 IMTFKDY 12

```

RESULT 12  
AAM90337  
ID AAM90337 standard; Protein: 57 AA.  
XX  
XX AAM90337;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen SEQ ID NO:17930.  
XX  
XX Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
XX

PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234222.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235833.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237044.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241788.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246539.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM,  
PI WPI: 2001-483426/52.  
XX N-PSDB; AAK63118.  
DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
PS Claim 11: SEQ ID NO 17930; 3071bp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169

CC represent sequences used in the exemplification of the present invention.  
XX SQ Sequence 57 AA;  
Query Match 3.5%; Score 7; DB 22; Length 57;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 38 SATSFTSL 44  
Db 50 SATSFTSL 56  
RESULT 13  
AAM82669  
ID AAM82669 standard; Protein: 80 AA.  
AC AAM82669;  
XX  
XX 07-NOV-2001 (first entry)  
DT  
XX  
DE Human Immune/haematopoietic antigen SEQ ID NO:10262.  
XX  
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW Cytostatic; gene therapy; vaccine; metastasis.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 16-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI: 2001-483426/52.  
 DR N-PSDB; AAK53450.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PT  
 PS Claim 11: SEQ ID NO 10262; 3071pp + Sequence Listing; English.  
 XX  
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC to AAK64703  
 CC to AAK67694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 80 AA:  
 Query Match: 3.5%; Score 7; DB 22; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 18-AUG-1998 (first entry)  
 XX  
 DE Actinomadura hibisca polyketide synthase protein 11.  
 XX  
 XX Miltienzyme; infection; fungi; yeast; gram-positive bacteria; virus;  
 KW dihydrobenzo(a)naphthacenequinone aglycon; antibiotic; pradimicin.  
 XX  
 OS Actinomadura hibisca.  
 XX  
 PN WO9811230-A1.  
 XX  
 PD 19-MAR-1998.  
 XX  
 PF 13-SEP-1996; 96WO-US14791.  
 XX  
 PR 13-SEP-1996; 96WO-US14791.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Dairi T, Oki T;  
 XX  
 DR WPI: 1998-207391/18.  
 DR N-PSDB; AAV26509.  
 XX  
 PT Actinomadura polyketide synthase genes - useful for preparation of  
 PT pradimicin  
 XX  
 PS Disclosure; Page 57; 71pp; English.  
 XX  
 CC The Actinomadura hibisca polyketide synthase proteins AAM54380-W84390  
 CC form a multienzyme complex. The enzyme is used for the biosynthesis of  
 CC a dihydrobenzo(a)naphthacenequinone aglycon preferably a pradimicin  
 CC which is an antibiotic useful against systemic fungal infections caused  
 CC by Candida albicans, Aspergillus fumigatus and Cryptococcus neoformans.  
 CC It is also active against a wide variety of fungi and yeasts, some  
 CC Gram-positive bacteria and viruses.  
 CC  
 XX  
 SQ Sequence 114 AA:  
 Query Match: 3.5%; Score 7; DB 19; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 VRUSROL 176  
 DB 73 VRUSROL 79

RESULT 15  
 ID AAV31332  
 XX  
 AC AAV31332;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secreted protein #1823.  
 XX  
 KW Human; vaccination; gene therapy; nutritional supplement;  
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200179449-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 16-APR-2001; 2001WO-US08656.  
 XX  
 PR 18-APR-2000; 2000US-0552929.  
 PR 26-JAN-2001; 2001US-0770160.  
 XX

PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy -  
 XX

PS Claim 20; Page 426; 765pp; English.

XX  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemias. AA029510-AA03304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.  
 XX

SQ Sequence 131 AA;

Query Match 3.5%; Score 7; DB 22; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GLRRLHR 131  
 |||||  
 Db 119 GLRRLHR 125

Search completed: July 7, 2003, 23:43:14  
 Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 23:43:17 ; Search time 24 Seconds  
(without alignments)  
948,692 Million cell updates/sec

Title: US-09-966-880a-8

Perfect score: 198  
Sequence: 1 MDSLIMNRKFLYQFKNRYM.....ILPLFYEDLDLRAPRTGL 198

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 440663 seqs, 114992915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 440663

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications\_AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCF\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCFUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	198	100.0	9	US-09-966-880a-8	Sequence 8, Appli
2	42	21.2	198	US-09-966-880a-2	Sequence 2, Appli
3	7	3.5	54	US-09-864-761-38108	Sequence 38108, A
4	7	3.5	222	US-09-925-300-1839	Sequence 1839, Ap
5	7	3.5	327	US-10-156-761-10683	Sequence 10683, A
6	7	3.5	350	US-09-881-752A-234	Sequence 234, App
7	7	3.5	384	US-09-729-674-114	Sequence 174, App
8	7	3.5	513	US-09-738-626-6336	Sequence 6326, Ap
9	7	3.5	932	US-10-079-429-4	Sequence 4, Appli
10	7	3.5	932	US-09-812-697-6	Sequence 6, Appli
11	7	3.5	932	US-09-912-697-8	Sequence 8, Appli
12	7	3.5	932	US-09-760-285-18	Sequence 18, Appli
13	7	3.5	932	US-09-788-657-17	Sequence 17, Appli
14	7	3.5	932	US-09-788-657-18	Sequence 18, Appli
15	7	3.5	944	US-10-213-990-27	Sequence 27, Appli
16	7	3.5	1178	US-10-041-856-9	Sequence 9, Appli
17	6	3.0	12	US-09-801-185A-31	Sequence 31, Appli
18	6	3.0	50	US-09-764-877-1554	Sequence 1554, Ap
19	6	3.0	54	US-09-864-761-46734	Sequence 46734, A

20	6	3.0	57	10	US-09-864-761-46082	Sequence 46082, A
21	6	3.0	61	10	US-09-867-550-2048	Sequence 2048, Ap
22	6	3.0	65	10	US-09-864-761-44925	Sequence 44925, A
23	6	3.0	81	8	US-08-424-550B-187	Sequence 187, App
24	6	3.0	82	10	US-09-764-877-1582	Sequence 1582, Ap
25	6	3.0	83	9	US-10-083-357-968	Sequence 968, App
26	6	3.0	85	9	US-10-156-761-12800	Sequence 12800, A
27	6	3.0	98	10	US-09-864-761-34110	Sequence 34110, A
28	6	3.0	98	10	US-09-864-761-34154	Sequence 34154, A
29	6	3.0	101	9	US-10-004-381-12	Sequence 12, Appli
30	6	3.0	101	9	US-10-156-761-7988	Sequence 7988, Ap
31	6	3.0	108	9	US-10-193-653-32	Sequence 32, Appli
32	6	3.0	111	10	US-09-864-761-35163	Sequence 35163, A
33	6	3.0	115	9	US-09-860-670-90	Sequence 90, Appli
34	6	3.0	119	9	US-09-764-891-3996	Sequence 3996, Ap
35	6	3.0	120	10	US-09-864-761-37707	Sequence 37707, A
36	6	3.0	156	9	US-09-981-876-136	Sequence 136, App
37	6	3.0	156	9	US-09-148-545-135	Sequence 135, App
38	6	3.0	159	10	US-09-867-550-954	Sequence 954, App
39	6	3.0	170	9	US-10-083-720A-4	Sequence 4, Appli
40	6	3.0	170	10	US-09-363-993-4	Sequence 19, Appli
41	6	3.0	170	10	US-09-916-790-19	Sequence 19, Appli
42	6	3.0	172	9	US-09-975-719-71	Sequence 51, Appli
43	6	3.0	178	9	US-09-965-528-55	Sequence 55, Appli
44	6	3.0	178	9	US-10-050-552A-1	Sequence 1, Appli
45	6	3.0	178	9	US-10-083-720A-6	Sequence 6, Appli

## ALIGNMENTS

RESULT 1  
US-09-966-880a-8  
Sequence 8, Application US/09966880A  
Patent No. US2002016473A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT FILING DATE: 2001-09-28  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: PCF/JPO0/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
TYPE: PRT  
LENGTH: 198  
ORGANISM: Homo sapiens  
US-09-966-880a-8  
Query Match 100.0%; Score 198; DB 9; Length 198;  
Best Local Similarity 100.0%; Pred. No. 1.5e-183;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MDSLIMNRKFLYQFKNRYAKGRREYLLCYVARRDSATSPSLDFGLRNKNCHEVELL 60  
DB 1 MDSLIMNRKFLYQFKNRYAKGRREYLLCYVARRDSATSPSLDFGLRNKNCHEVELL 60  
QY 61 FLRTISDMDLDPGRCYRYTWTFTSMSPCYDCARHAYADLRGNPNLSLITFARLYFCDCDR 120  
DB 61 FLRTISDMDLDPGRCYRYTWTFTSMSPCYDCARHAYADLRGNPNLSLITFARLYFCDCDR 120  
QY 121 AEPGRLRLHAGVOIALMFKDFYCWNTFVENHETFTFAMEGLHENSRYLSROLRRIL 180  
DB 121 AEPGRLRLHAGVOIALMFKDFYCWNTFVENHETFTFAMEGLHENSRYLSROLRRIL 180

OY 181 LPLEYVDLRLDAFRTGL 198  
|||||  
Db 181 LPLEYVDLRLDAFRTGL 198

RESULT 2  
US-09-966-880A-2  
Sequence 2, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 198  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-966-880A-2

Query Match 21.2%; Score 42; DB 9; Length 198;  
Best Local Similarity 100.0%; Pred. No. 8,7e-33;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 GCHVELLFLRISDWDLPGRCRYRTWTSPCYDCARHYA 95  
|||||  
Db 54 GCHVELLFLRISDWDLPGRCRYRTWTSPCYDCARHYA 95

RESULT 3  
US-09-864-761-38108  
Sequence 38108, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38108  
LENGTH: 54  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000498.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.97  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2  
OTHER INFORMATION: EST\_HUMAN HIT: AA121545.1, EVALUATE 1.00e-03  
US-09-864-761-38108

Query Match 3.5%; Score 7; DB 10; Length 54;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 138 IMTFKDY 144  
|||||  
Db 6 IMTFKDY 12

RESULT 4  
US-09-925-300-1639  
Sequence 1639, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1639  
LENGTH: 222  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-300-1639

Query Match 3.5%; Score 7; DB 10; Length 222;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 108 IFTARLY 114  
|||||  
DB 150 IFTARLY 156

RESULT 5  
US-10-156-761-10683  
; Sequence 10683, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10683  
; LENGTH: 327  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10683

Query Match 3.5%; Score 7; DB 9; Length 327;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 PLYEVD 188  
|||||  
DB 303 PLYEVD 309

RESULT 6  
US-09-881-752A-234  
; Sequence 234, Application US/09881752A  
; Patent No. US20020115078A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleantchous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Comen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the  
; FILE REFERENCE: 06132/041002  
; CURRENT APPLICATION NUMBER: US/09/881,752A  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 08/833,457  
; PRIOR FILING DATE: 1997-04-01  
; NUMBER OF SEQ ID NOS: 370  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 234  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-881-752A-234

Query Match 3.5%; Score 7; DB 10; Length 350;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 DFLRGNP 102  
|||||

DB 209 DFLRGNP 215

RESULT 7  
US-09-729-674-174  
; Sequence 174, Application US/09729674  
; Patent No. US2001003935A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Treacy, David  
; APPLICANT: Merberg, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6055-64X  
; CURRENT APPLICATION NUMBER: US/09/729,674  
; CURRENT FILING DATE: 2000-12-04  
; PRIOR APPLICATION NUMBER: 09/539,330  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 283  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 174  
; LENGTH: 384  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-729-674-174

Query Match 3.5%; Score 7; DB 10; Length 384;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 FTSMSPC 87  
|||||  
DB 282 FTSMSPC 288

RESULT 8  
US-09-738-626-6326  
; Sequence 6326, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: IATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 6326

```

; LENGTH: 513
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6326

Query Match          3.5%; Score 7; DB 9; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      38 SATSFSL 44
      |||||
DB      322 SATSFSL 328

RESULT 9
US-10-079-429-4
; Sequence 4, Application US/10079429
; Publication No. US20030027177A1
; GENERAL INFORMATION:
; APPLICANT: Haseltine et al.
; TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
; FILE REFERENCE: PFI06P3D1
; CURRENT APPLICATION NUMBER: US/10/079,429
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: PCT/US95/01035
; PRIOR FILING DATE: 1995-01-25
; PRIOR APPLICATION NUMBER: 08/468,024
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/465,769
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/294,312
; PRIOR FILING DATE: 1994-08-23
; PRIOR APPLICATION NUMBER: 08/210,143
; PRIOR FILING DATE: 1994-03-16
; PRIOR APPLICATION NUMBER: 08/187,757
; PRIOR FILING DATE: 1994-01-27
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 932
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-079-429-4

Query Match          3.5%; Score 7; DB 9; Length 932;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      170 VRLSROL 176
      |||||
DB      882 VRLSROL 888

RESULT 10
US-09-912-697-6
; Sequence 5, Application US/09912697
; Publication No. US20030068808A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas C
; APPLICANT: Sassi, Philip M
; APPLICANT: Grasso, Luigi M
; APPLICANT: Kline, J Bradford
; TITLE OF INVENTION: METHODS FOR GENERATING ANTIBIOTIC RESISTANT MICROBES AND NOVEL
; FILE REFERENCE: MOR-0040
; CURRENT APPLICATION NUMBER: US/09/912,697
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-912-697-6

Query Match          3.5%; Score 7; DB 9; Length 932;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      170 VRLSROL 176
      |||||
DB      882 VRLSROL 888

RESULT 11
US-09-912-697-8
; Sequence 8, Application US/09912697
; Publication No. US20030068808A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas C
; APPLICANT: Sassi, Philip M
; APPLICANT: Grasso, Luigi M
; APPLICANT: Kline, J Bradford
; TITLE OF INVENTION: METHODS FOR GENERATING ANTIBIOTIC RESISTANT MICROBES AND NOVEL
; FILE REFERENCE: MOR-0040
; CURRENT APPLICATION NUMBER: US/09/912,697
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-912-697-8

Query Match          3.5%; Score 7; DB 9; Length 932;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      170 VRLSROL 176
      |||||
DB      882 VRLSROL 888

RESULT 12
US-09-760-285-18
; Sequence 18, Application US/09760285
; Publication No. US20030091997A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas C
; APPLICANT: Grasso, Luigi M
; APPLICANT: Sassi, Philip M
; TITLE OF INVENTION: CHEMICAL INHIBITORS OF MISMATCH REPAIR
; FILE REFERENCE: MOR-0017
; CURRENT APPLICATION NUMBER: US/09/760,285
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-760-285-18

Query Match          3.5%; Score 7; DB 9; Length 932;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      170 VRLSROL 176
      |||||
DB      882 VRLSROL 888

RESULT 13
US-09-788-657-17
```

```
; Sequence 17, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sassi, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/09/788,657
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-788-657-17
```

```
Query Match          3.5%; Score 7; DB 10; Length 932;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      170 VRLSROL 176
        |||||||
Db      882 VRLSROL 888
```

```
RESULT 14
US-09-788-657-18
; Sequence 18, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sassi, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/09/788,657
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-788-657-18
```

```
Query Match          3.5%; Score 7; DB 10; Length 932;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      170 VRLSROL 176
        |||||||
Db      882 VRLSROL 888
```

```
RESULT 15
US-10-213-990-27
; Sequence 27, Application US/10213990
; Publication No. US20030082595A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Bussey, Howard
```

```
; APPLICANT: Storms, Reg
; APPLICANT: Roemer, Terry
; TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
; FILE REFERENCE: 10182-019-999
; CURRENT APPLICATION NUMBER: US/10/213,990
; CURRENT FILING DATE: 2002-08-05
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Aspergillus
US-10-213-990-27
```

```
Query Match          3.5%; Score 7; DB 9; Length 944;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      39 ATSFSLD 45
        |||||||
Db      893 ATSFSLD 899
```

```
Search completed: July 7, 2003, 23:46:08
Job time : 24 secs
```

TELEFAX: (310) 447-4300  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
US-08-591-438-20

Query Match 3.0%; Score 6; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 DLRDAF 193  
|||||  
5-DIPAD 10

Mon Jul 14 14:47:06 2003

us-09-966-880a-8.olig.ra1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 23:42:32; Search time 20 Seconds  
(Without alignments)  
291.287 Million cell updates/sec

Title: US-09-966-880a-8  
Perfect score: 198  
Sequence: 1 MDSLIMRKRFLYQFKVNR.....ILLPLYEVDLDRDAFRTGL 198

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size: 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	3.5	190	1	US-08-816-241-1
2	7	3.5	190	3	US-09-128-395-1
3	7	3.5	932	4	US-08-294-312B-4
4	7	3.5	932	4	US-08-294-312B-4
5	6	3.0	11	2	US-08-468-024B-4
6	6	3.0	11	2	US-08-591-438-20
7	6	3.0	12	1	US-07-778-233B-31
8	6	3.0	12	1	US-08-290-641-31
9	6	3.0	12	1	US-08-548-540-31
10	6	3.0	12	3	US-08-599-226-31
11	6	3.0	12	4	US-09-125-098-31
12	6	3.0	12	5	PCT-US96-09809-31

RESULT 7  
US-07-963-321-31  
Sequence 31, Application US/07963321  
Patent No. 533865  
GENERAL INFORMATION:  
APPLICANT: Schatz, Peter J.  
APPLICANT: Miller, Jeff F.  
APPLICANT: Stemmer, Willem P.C.  
TITLE OF INVENTION: Peptide Library and Screening Method  
NUMBER OF SEQUENCES: 91  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: William M. Smith  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA

28	6	3.0	92	4	US-09-394-630-2	Sequence 2, Appl
29	6	3.0	108	4	US-09-113-977C-32	Sequence 32, Appl
30	6	3.0	116	1	US-08-687-895-3	Sequence 3, Appl
31	6	3.0	116	1	US-08-816-241-3	Sequence 3, Appl
32	6	3.0	116	2	US-09-040-482-3	Sequence 3, Appl
33	6	3.0	116	3	US-09-128-395-3	Sequence 3, Appl
34	6	3.0	116	4	US-09-201-227A-17	Sequence 17, Appl
35	6	3.0	120	4	US-09-201-227A-2	Sequence 2, Appl
36	6	3.0	147	1	US-08-270-805C-4	Sequence 4, Appl
37	6	3.0	147	2	US-08-410-654B-4	Sequence 4, Appl
38	6	3.0	147	2	US-08-474-851-4	Sequence 4, Appl
39	6	3.0	147	2	US-08-481-560-4	Sequence 4, Appl
40	6	3.0	147	3	US-08-170-113-4	Sequence 4, Appl
41	6	3.0	147	4	US-08-643-810A-4	Sequence 4, Appl
42	6	3.0	147	4	US-09-552-613-4	Sequence 4, Appl
43	6	3.0	147	5	PCT-US93-07646-2	Sequence 2, Appl
44	6	3.0	154	4	US-09-452-624A-3	Sequence 3, Appl
45	6	3.0	156	5	PCT-US94-08052-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-816-241-1  
Sequence 1, Application US/08816241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,241  
CLASSIFICATION: 435  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:

Mon Jul 14 14:47:06 2003

us-09-966-880a-8.olig.ra1

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 170 VRLSROL 176  
Db 882 VRLSROL 888

RESULT 4  
US-08-468-024B-4  
Sequence 4, Application US/08468024B  
Patent No. 6416984  
GENERAL INFORMATION:  
APPLICANT: Haseltine et al.  
TITLE OF INVENTION: Human DNA Mismatch Repair Proteins  
FILE REFERENCE: PFI06P3  
CURRENT APPLICATION NUMBER: US/08/468,024B  
CURRENT FILING DATE: 1995-06-06  
PRIOR APPLICATION NUMBER: 08/294,312

APPLICANT: Cull, Willard G.  
APPLICANT: Miller, Jeff F.  
APPLICANT: Stemmer, William P.C.  
TITLE OF INVENTION: Peptide Library and Screening Method  
NUMBER OF SEQUENCES: 91  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: William M. Smith  
STREET: One Market Plaza, Stewart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,641  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,321  
FILING DATE: 15-OCT-1992  
APPLICATION NUMBER: US 07/778,223  
FILING DATE: 15-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11509-50-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
CLONE: 56 3 1.1  
US-08-290-641-31

Query Match 3.0%; Score 6; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LRLRLR 131  
Db 1 LRLRLR 6

RESULT 9  
US-08-548-540-31  
Sequence 31, Application US/08548540  
Patent No. 5733731  
GENERAL INFORMATION:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/548,540  
FILING DATE: 26-OCT-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/290,641  
FILING DATE: 15-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,321  
FILING DATE: 15-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 16528J-001240US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
CLONE: 56 3 1.1  
US-08-548-540-31

Query Match 3.0%; Score 6; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LRLRLR 131  
Db 1 LRLRLR 6

RESULT 10  
US-08-599-226-31  
Sequence 31, Application US/08599226  
Patent No. 6090382  
GENERAL INFORMATION:  
APPLICANT: Salfeld, Jochen G.  
APPLICANT: Allen, Deborah J.  
APPLICANT: Hoogenboom, Hendricus R.J.M.  
APPLICANT: Kaymakcalan, Zehra  
APPLICANT: Labkovsky, Boris  
APPLICANT: Mankovich, John A.  
APPLICANT: McGuinness, Brian T.  
APPLICANT: Roberts, Andrew J.  
APPLICANT: Sakorafas, Paul  
APPLICANT: Schoenhaut, David  
APPLICANT: Vaughan, Tristian J.  
APPLICANT: White, Michael

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/599,226  
FILING DATE: 08-FEB-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Decortt, Giulio A., Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: BBI-043  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-599-226-31

Query Match 3.0%; Score 6; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 TSFSLD 45  
Db 6 TSFSLD 11

RESULT 11  
US-09-125-098-31  
Sequence 31, Application US/09125098  
Patent No. 6258562  
GENERAL INFORMATION:  
APPLICANT: Safield, Jochen G.  
APPLICANT: Allen, Deborah J.  
APPLICANT: Hoogenboom, Hendricus R.J.M.  
APPLICANT: Kaymakalan, Zehra  
APPLICANT: Labkovsky, Boris  
APPLICANT: Markovich, John A.  
APPLICANT: McGuinness, Brian T.  
APPLICANT: Roberts, Andrew J.  
APPLICANT: Sakorafas, Paul  
APPLICANT: Schoenhaut, David  
APPLICANT: Vaughan, Trestan J.  
APPLICANT: White, Michael  
APPLICANT: Wilton, Andrew J.  
TITLE OF INVENTION: Human Antibodies that Bind Human TNFA  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/125,098  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/599,226  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Decortt, Giulio A., Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: BBI-043  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-09-125-098-31

Query Match 3.0%; Score 6; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 TSFSLD 45  
Db 6 TSFSLD 11

RESULT 12  
PCT-US96-09809-31  
Sequence 31, Application PC/TUS9609809  
GENERAL INFORMATION:  
APPLICANT: Schatz, Peter J.  
APPLICANT: Cull, Millard G.  
APPLICANT: Miller, Jeff F.  
APPLICANT: Stemmer, Willem P.C.  
APPLICANT: Gates, Christian M.  
TITLE OF INVENTION: Peptide Library and Screening Method  
NUMBER OF SEQUENCES: 162  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: William M. Smith  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/09809  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/548,540  
FILING DATE: 26-OCT-1995  
APPLICATION NUMBER: US 08/290,641  
FILING DATE: 15-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,321  
FILING DATE: 15-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 16528J-001240US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
CLONE: 56 3 1.1  
PCT-US96-09809-31

Query Match 3.0%; Score 6; DB 5; Length 12;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 LRLRLR 131  
DB 1 LRLRLR 6

## RESULT 13

US-08-280-157A-1  
Sequence 1, Application US/08280157A  
Patent No. 5648339  
GENERAL INFORMATION:  
APPLICANT: BINIE V. LIPPS  
TITLE OF INVENTION: HERPOXIN: HERPES VIRUS  
TITLE OF INVENTION: INHIBITOR AND METHOD  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BINIE V. LIPPS  
STREET: 4509 MIMOSA DR.  
CITY: BELLAIRE  
STATE: TEXAS  
COUNTRY: USA  
ZIP: 77401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISK, 1.2 MB  
COMPUTER: IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS 5.0/WINDOWS 3.1  
SOFTWARE: MS WORD 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/280,157A  
FILING DATE: 25 JULY 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: JOHN R. CASPERSON  
REGISTRATION NUMBER: 28,198  
REFERENCE/DOCKET NUMBER: FWL-PAT-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-723-6845  
TELEFAX: 713-663-7290  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15  
TYPE: AMINO ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN IN SEQ ID NO: 1  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N  
ORIGINAL SOURCE: SNAKE VENOM: SEQ ID NO: 1:  
ORGANISM: Naja N. KAOUTHIA  
STRAIN: WILD  
INDIVIDUAL ISOLATE: THAILAND WILD  
DEVELOPMENTAL STAGE: ADULT  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
ORGANELLE:  
IMMEDIATE SOURCE: SNAKE VENOM SEQ ID NO: 1:  
LIBRARY:  
CLONE:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:

ISSUE:  
PAGES:  
DATE:  
US-08-280-157A-1

Query Match 3.0%; Score 6; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LYOFKN 17  
DB 2 LYOFKN 7

## RESULT 14

US-08-053-131-80  
Sequence 80, Application US/08053131  
Patent No. 5661016  
GENERAL INFORMATION:  
APPLICANT: Lonberg, Nils  
APPLICANT: Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 5661016-Human Animals for  
TITLE OF INVENTION: Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 197  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 200  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/053,131  
FILING DATE: 26-APR-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,279  
FILING DATE: 17-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/853,408  
FILING DATE: 18-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 14643-9-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-053-131-80

Query Match 3.0%; Score 6; DB 1; Length 27;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 CARHYA 95  
DB 2 CARHYA 7

RESULT 15  
US-08-645-641-80  
Sequence 80, Application US/08645641  
Patent No. 5719032  
GENERAL INFORMATION:  
APPLICANT: Lombert, Nils  
APPLICANT: Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 5719032-Human Animals for  
TITLE OF INVENTION: Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 150  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: William M. Smith  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/645,641  
FILING DATE: 20-MAY-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/904,068  
FILING DATE: 23-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 14643-000913  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-645-641-80

## Query Match

3.0%; Score 6; DB 1; Length 27;

Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 CARVA 95  
|||||  
Db 2 CARVA 7

Search completed: July 7, 2003, 23:45:33  
Job time : 22 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 23:36:52; Search time 17 Seconds

(without alignments)  
483.077 Million cell updates/sec

Title: US-09-966-880a-8

Perfect score: 198

Sequence: 1 MDSLNNRRKFLYQFNKRW.....ILLPLYVDLDRDFTLGL 198

Scoring table:

Gapop 60.0, Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	8	4.0	695 1	NU5C_CAPBA
2	8	4.0	695 1	NU5C_LYCES
3	8	4.0	740 1	NU5C_TOBAC
4	8	4.0	1178 1	MNN4_YEAST
5	7	3.5	143 1	Y880_MYCTU
6	7	3.5	154 1	SM20_SCMA
7	7	3.5	521 1	CP11_LIMI
8	7	3.5	521 1	CP11_PLEPL
9	7	3.5	932 1	FMS1_HUMAN
10	7	3.5	1294 1	YA3B_SCHPO
11	6	3.0	16 1	PA2_NAUSP
12	6	3.0	27 1	PA21_MITNI
13	6	3.0	28 1	PA23_MITNI
14	6	3.0	97 1	LTUB_CHLPN
15	6	3.0	113 1	RPOA_BUCAP
16	6	3.0	117 1	PA2A_PSEPO
17	6	3.0	118 1	PA21_NAJME
18	6	3.0	118 1	PA21_NAJMO
19	6	3.0	118 1	PA22_NAJMO
20	6	3.0	118 1	PA23_NAJMO
21	6	3.0	118 1	PA23_NAJNG
22	6	3.0	118 1	PA26_BUNFA
23	6	3.0	118 1	PA2X_BUNFA
24	6	3.0	118 1	PA2_NAJRA
25	6	3.0	119 1	PA21_HEMHA
26	6	3.0	119 1	PA21_NAOX
27	6	3.0	119 1	PA22_ASPSC
28	6	3.0	119 1	PA22_NAJKA
29	6	3.0	119 1	PA22_NAJME
30	6	3.0	119 1	PA23_NAJKA
31	6	3.0	119 1	PA23_NAJME
32	6	3.0	119 1	PA2_NAJNA
33	6	3.0	126 1	INL3_DROME

34	6	3.0	131 1	RS9_HAINT	O9h9j2 halobacteri
35	6	3.0	135 1	YGRX_ECOLI	Q46824 escherichia
36	6	3.0	145 1	PA20_BUNMU	P00606 bungarus mu
37	6	3.0	146 1	PA2_NAJAT	P00598 najata atra (
38	6	3.0	150 1	DTD_YEAST	Q07648 saccharomyc
39	6	3.0	153 1	G8A6_DICDI	P34044 dictyostell
40	6	3.0	162 1	CAB2_BOVIN	Q9n199 bos taurus
41	6	3.0	162 1	Y157_AOUAE	O66547 aquilex aeo
42	6	3.0	170 1	BCRF_EBV	P03180 Epstein-Bar
43	6	3.0	170 1	E1B5_ADEC2	P35983 canine aden
44	6	3.0	174 1	GNTR_ECOLI	P46859 escherichia
45	6	3.0	176 1	RL10_SUITO	Q96y44 sulfolobus

## ALIGNMENTS

RESULT 1	NU5C_CAPBA	STANDARD;	PRT;	695 AA.
ID	NU5C_CAPBA			
AC	Q31952;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	NADH-plastoquinone oxidoreductase chain 5, chloroplast (EC 1.6.5.3)			
GN	(Fragment).			
OS	Capsicum baccatum.			
OC	Chloroplast.			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
OC	Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.			
OX	NCBI_TaxID=33114;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Olmstead R.G., Sweere J.A.;			
RT	"Combining data in phylogenetic systematics: an empirical approach			
RT	using three molecular data sets in the Solanaceae.";			
RL	Syst. Biol. 43:467-481(1994).			
CC	-1- CATALYTIC ACTIVITY: NADH + plastoquinone -> NAD(+) + plastoquinol.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U08916; AAA18598.1; -			
DR	InterPro; IPR001750; Oxidored_q1.			
DR	InterPro; IPR002128; Oxidored_q1_C.			
DR	InterPro; IPR001516; Oxidored_q1_N.			
DR	Pfam; PF00361; Oxidored_q1; 1.			
DR	Pfam; PF00662; Oxidored_q1_N; 1.			
DR	Pfam; PF01010; Oxidored_q1_C; 1.			
KW	Oxidoreductase; NAD; Plastoquinone; Chloroplast.			
FT	NON_TER 1			
FT	NON_TER 695			
FT	SEQUENCE 695 AA; 78709 MW; 80D56D1C47338C57 CRC64;			
QY	42 FSLDFGYL 49			
QY				
Db	73 FSLDFGYL 80			
RESULT 2	NU5C_LYCES	STANDARD;	PRT;	695 AA.
ID	NU5C_LYCES			
Query Match	4.0%; Score 8; DB 1; Length 695;			
Best Local Similarity	100.0%; Pred. No. 2.2;			
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			

```

AC Q32516;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NADH-plastoquinone oxidoreductase chain 5, chloroplast (EC 1.6.5.3)
DE (Fragment).
GN NDHF.
OS Lycopersicon esculentum (Tomato).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RA Olmstead R.G., Sweere J.A.;
RL Submitted (May-1994) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + plastoquinone = NAD(+) + plastoquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 008921; AAA18603.1; -
CC InterPro: IPR001750; Oxidored_q1.
CC InterPro: IPR002128; Oxidored_q1_C.
CC InterPro: IPR001516; Oxidored_q1_N.
CC Pfam; PF00361; Oxidored_q1; 1.
CC Pfam; PF00662; Oxidored_q1_N; 1.
CC Pfam; PF01010; Oxidored_q1_C; 1.
CC Oxidoreductase; NAD: Plastoquinone; Chloroplast.
CC NON_TER 1
CC FT NON_TER 1
CC FT SEQUENCE 695 AA; 78386 MW; 639B0F4C0778DEA9 CRC64;
SQ
Query Match 4.0%; Score 8; DB 1; Length 695;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 42 FSLDFGYL 49
DB 73 FSLDFGYL 80
RESULT 3
ID NU5C_TOBAC STANDARD; PRT; 740 AA.
AC P06265;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUN-1984 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-plastoquinone oxidoreductase chain 5, chloroplast (EC 1.6.5.3).
GN NDHF OR NDH5.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv Bright Yellow 4;
RA Shinozaki K., Ome M., Tanaka M., Wakasugi T., Hayashida N.,
RA Matsuyashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.-Y., Sugita M.,
RA Dero H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA Todoroh N., Shimada H., Sugiyama M.;
RA "The complete nucleotide sequence of the tobacco chloroplast genome:
RA its gene organization and expression."

```

```

RL EMBO J. 5:2043-2049(1986).
RN [2]
RP REVISIONS.
RX MEDLINE=94003079; PubMed=8400137;
RA Olmstead R.G., Sweere J.A., Wolfe K.H.;
RT "Ninety extra nucleotide in ndhf gene of tobacco chloroplast DNA: a
RT summary of revisions to the 1986 genome sequence.";
RL Plant Mol. Biol. 22:1191-1193(1993).
CC -1- CATALYTIC ACTIVITY: NADH + plastoquinone = NAD(+) + plastoquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14953; AAA84685.1; -
CC EMBL; Z00044; CAA77430.1; -
CC PIR; A00454; DEN1M5.
CC PIR; S37352; S37352.
CC InterPro: IPR003916; NADHoxred5.
CC InterPro: IPR001750; Oxidored_q1.
CC InterPro: IPR002128; Oxidored_q1_C.
CC InterPro: IPR001516; Oxidored_q1_N.
CC Pfam; PF00361; Oxidored_q1; 1.
CC Pfam; PF00662; Oxidored_q1_N; 1.
CC Pfam; PF01010; Oxidored_q1_C; 1.
CC PRINTS; PR01434; NADHDMNASE5.
CC Oxidoreductase; NAD: Plastoquinone; Chloroplast.
CC SEQUENCE 740 AA; 83717 MW; D6515B4B8BF0B54 CRC64;
SQ
Query Match 4.0%; Score 8; DB 1; Length 740;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 42 FSLDFGYL 49
DB 81 FSLDFGYL 88
RESULT 4
ID MNN4_YEAST STANDARD; PRT; 1178 AA.
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1996 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MNN4 protein.
GN MNN4 OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=S288C;
RX MEDLINE=97175967; PubMed=9023541;
RA Odani T., Shima Y.-T., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MNN4 gene required for phosphorylation
RT of N-linked oligosaccharides in Saccharomyces cerevisiae.";
RL Glycobiology 6:805-810(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (Mar-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
CC MANNOSYLPHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
CC MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
CC PORTIONS OF N-LINKED OLIGOSACCHARIDES.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).

```

CC	-1	SIMILARITY: TO YEAST YJ061W
CC	-1	CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC	-1	FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC		or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC		-----
DR	EMBL; D83006; BAA11676.1; -	
DR	EMBL; Z28201; CAA82046.1; -	
DR	EMBL; Z28200; CAA82044.1; -	
DR	PIR; S38037; S38037.	
DR	PIR; S38038; S38038.	
DR	SGD; S0001683; MNNA.	
KW	Transmembrane; Signal-anchor; Repeat.	
FT	DOMAIN 1 27	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 28 48	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT		(POTENTIAL).
FT	DOMAIN 49 1178	LUMENAL (POTENTIAL).
FT	DOMAIN 1032 1174	ARG/GLU/LYS-RICH (HIGHLY CHARGED).
FT	DOMAIN 1042 1174	17 X 8 AA TANDEM REPEATS OF K-K-K-E-E-
FT		E-E.
FT	REPEAT 1042 1049	1.
FT	REPEAT 1050 1057	2.
FT	REPEAT 1058 1065	3.
FT	REPEAT 1066 1073	4.
FT	REPEAT 1074 1081	5.
FT	REPEAT 1082 1089	6.
FT	REPEAT 1090 1097	7 (APPROXIMATE).
FT	REPEAT 1098 1105	8.
FT	REPEAT 1106 1113	9 (APPROXIMATE).
FT	REPEAT 1114 1121	10 (APPROXIMATE).
FT	REPEAT 1122 1129	11 (APPROXIMATE).
FT	REPEAT 1130 1137	12.
FT	REPEAT 1138 1144	13 (APPROXIMATE).
FT	REPEAT 1145 1152	14 (APPROXIMATE).
FT	REPEAT 1153 1160	15 (APPROXIMATE).
FT	REPEAT 1161 1168	16 (APPROXIMATE).
FT	REPEAT 1169 1174	17 (INCOMPLETE).
FT	DOMAIN 37 40	POLY-ILE.
FT	DOMAIN 1042 1045	POLY-LYS.
FT	DOMAIN 1046 1049	POLY-GLU.
FT	DOMAIN 1050 1053	POLY-LYS.
FT	DOMAIN 1054 1057	POLY-GLU.
FT	DOMAIN 1058 1061	POLY-LYS.
FT	DOMAIN 1062 1065	POLY-GLU.
FT	DOMAIN 1066 1069	POLY-LYS.
FT	DOMAIN 1070 1073	POLY-GLU.
FT	DOMAIN 1074 1077	POLY-LYS.
FT	DOMAIN 1078 1081	POLY-GLU.
FT	DOMAIN 1082 1085	POLY-LYS.
FT	DOMAIN 1086 1089	POLY-GLU.
FT	DOMAIN 1094 1097	POLY-LYS.
FT	DOMAIN 1098 1101	POLY-LYS.
FT	DOMAIN 1102 1105	POLY-GLU.
FT	DOMAIN 1134 1137	POLY-GLU.
FT	DOMAIN 1157 1160	POLY-GLU.
FT	DOMAIN 1165 1168	POLY-GLU.
SQ	SEQUENCE 1178 AA; 139360 MM; BC05DAEDAEFCB282 CRC64;	
QY	Query Match	4.0%; Score 8; DB 1; Length 1178;
	Best Local Similarity	100.0%; Pred. No. 3.6;
	Matches 8; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
DB	176 LRRILPL 183	
	25 LRRILPL 32	

```

RESULT 5
RESULT 6
EM20_SCHMA

AC ID Y880_MYCTU STANDARD; PRT: 143 AA.
AC Q10542;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transcriptional regulator RV0880.
GN RV0880 OR MT0903.1 OR MTCY31.08.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteriidae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1773;
[1]
XP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsley T., Jags K., Kitchell A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickley E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; Z73101; CAAG7387.1; -
CC DR EMBL; AE006977; -; NOT_ANNOTATED_CDS.
CC DR TIGR; MT0903.1; -
CC DR Tuberculist; RV0880; -
CC DR InterPro; IPR000835; HTH_Marr.
CC DR Pfam; PF01047; Marr; 1.
CC DR SMART; SM00347; HTH_MARR; 1.
CC DR PROSITE; PS01117; HTH_MARR_FAMILY; FALSE_NEG.
CC KW Hypothetical protein; Transcription regulation; DNA-binding;
CC KM Complete proteome.
CC SQ SEQUENCE 143 AA; 15576 MW; F9E2D59CD77E0CE8 CXC64;

Query Match 3.5%; Score 7; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 171 RLSQLR 177
1111111
DB 18 RLSQLR 24

```

```

ID SM20_SCHMA STANDARD; PRT; 154 AA.
AC P15845;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 20 kDa calcium-binding protein (Antigen SM20).
GN SM20.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidae;
OC Schistosomatoidea; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RA Stewart T.J., Smith A.L., Havercroft J.C.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 65-154 FROM N.A.
RC STRAIN=Puerto Rican;
RX MEDLINE=90220753; PubMed=2325706;
RA Havercroft J.C., Huggins M.C., Dunne D.W., Taylor D.W.;
RT "Characterisation of SM20, a 20-kilodalton calcium-binding protein of
RT Schistosoma mansoni."
RL Mol. Biochem. Parasitol. 38:211-220(1990).
CC -1- FUNCTION: CALCIUM-BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE TEGUMENTAL MEMBRANE.
CC -1- DEVELOPMENTAL STAGE: SM20 IS EXPRESSED IN SCHISTOSOMULA AND
CC ADULT WORMS, BUT NOT IN EGGS.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M94045; AAA29921.1; -
DR EMBL; M28888; AAA29926.1; -
DR PIR; A44977; A44977.
DR HSSP; P02593; ICDM.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 4.
DR ProDom; PD000012; EF-hand; 2.
DR SMART; SM00054; EFh; 3.
DR PROSITE; PS00018; EF_HAND_1.
KW Calcium-binding; Antigen; Repeat; Membrane.
FT DOMAIN 26 35 ANCESTRAL CALCIUM SITE 1 (POTENTIAL).
FT CA_BIND 62 73 EF-HAND 2 (POTENTIAL).
FT CA_BIND 99 110 EF-HAND 3 (POTENTIAL).
FT CA_BIND 135 146 EF-HAND 4 (POTENTIAL).
SQ SEQUENCE 154 AA; 17894 MW; 36A6694780BDAB2 CRC64;

Query Match 3.5%; Score 7; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 6; 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 DLRDAFR 194
DB 90 DLRDAFR 96

```

```

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Limanda.
OX NCBI_TaxID=27771;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=21315588; PubMed=11423384;
RA Craft J.A., Robertson F.E., McPhail M.E., Brown E., Stagg R.M.;
RT "Measurement of cytochrome P4501A induction in dab (Limanda limanda)
RT and other teleosts with species-specific cDNA probes: isolation and
RT characterisation of dab cDNA and its use in expression studies with
RT beta-naphthoflavone-treated fish."
RL Comp. Biochem. Physiol. 129C:115-127(2001).
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ001724; CAA04953.1; -
DR HSSP; P00179; LDR6.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT BINDING 463 463 HEME (BY SIMILARITY).
SQ SEQUENCE 521 AA; 59063 MW; 11DE326C8A7FBD99 CRC64;

Query Match 3.5%; Score 7; DB 1; Length 521;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 PEGJRL 129
DB 39 PEGJRL 45

```

```

RESULT 7
CPIL_LIMLI STANDARD; PRT; 521 AA.
ID CPIL_LIMLI
AC 042430;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 1A1 (EC 1.14.14.1) (CYP1A1).
GN CYP1A1.
OS Limanda limanda (Dab).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

RESULT 8
CPIL_PLEPL STANDARD; PRT; 521 AA.
ID CPIL_PLEPL
AC 092100;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P450 1A1 (EC 1.14.14.1) (CYP1A1).
GN CYP1A1.
OS Pleuronectes platessa (Plaice).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Pleuronectes.
OX NCBI_TaxID=8262;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94251288; PubMed=8193668;
RA Leaver M.J., Pirrit L., George S.G.;
RT "Cytochrome P450 1A1 cDNA from plaice (Pleuronectes platessa) and
RT induction of P450 1A1 mRNA in various tissues by 3-methylcholanthrene
RT and isosafrole."

```

```

RL Mol. Biol. Biotechnol. 2:338-345(1993).
CC -1- FUNCTION: CYTOCHROME P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- INDUCTION: BY 3-METHYLCOLANTHRENE (3MC) AND ISOSAFROLE (ISF).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73631; CA52010.1; -.
DR HSSP: P00179; IDT6.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT BINDING 463 463 HEME (BY SIMILARITY).
SQ SEQUENCE 521 AA; 59061 MW; AA7A97CB4578F9E6 CRC64;

Query Match 3.5%; Score 7; DB 1; Length 521;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 123 PEGLRRL 129
Db 39 PEGLRRL 45

RESULT 9
PMS1_HUMAN STANDARD; PRT; 932 AA.
AC P54277;
DR 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE PMS1 protein homolog 1 (DNA mismatch repair protein PMS1).
GN PMS1 OR PMSL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gal bladder;
RX MEDLINE=9435394; PubMed=8072530;
RA Nicotlaides N.C., Papadopoulos N., Liu B., Wei Y.-F., Carter K.C.,
RA Ruben S.M., Rosen C.A., Haseltine W.H., Fleischmann R.D.,
RA Fraser C.M., Adams M.D., Venter J.C., Dunlop M.G., Hamilton S.R.,
RA Petersen G.M., de la Chapelle A., Vogelstein B., Kinzler K.W.;
RT Mutations of two PMS homologues in hereditary nonpolyposis colon
RT cancer."
RL Nature 371:75-80(1994).
RN [2]
RP VARIANTS HNPCC THR-394 AND ARG-501.
RX MEDLINE=94408236; PubMed=10480359;
RA Wang Q., Lasset C., Desseigne F., Saurin J.-C., Maugard C.,
RA Navarro C., Ruano E., Descos L., Trillet-Lenoir V., Bosset J.-F.,
RA Puitsieux A.;
RT "Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and
RT hMSH6 genes in 75 French kindreds with nonpolyposis colorectal
RT cancer."
RL Hum. Genet. 105:79-85(1999).
CC -1- FUNCTION: PROBABLY INVOLVED IN THE REPAIR OF MISMATCHES IN DNA.
```

```

CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- DISEASE: ASSOCIATED WITH FAMILIAL HEREDITARY NONPOLYPOSIS COLON
CC CANCER (HNPCC) (LYNCH SYNDROME). HNPCC IS ONE OF THE MOST COMMON
CC GENETIC DISEASES IN THE WESTERN WORLD, AND ACCOUNTS FOR 1% OF ALL
CC COLON CANCERS.
CC -1- DISEASE: DEFECTS IN PMS1 ARE A CAUSE OF HEREDITARY NON-POLYPOSIS
CC COLORECTAL CANCER (HNPCC) (LYNCH SYNDROME). HNPCC IS AN AUTOSOMAL,
CC DOMINANTLY INHERITED DISEASE ASSOCIATED WITH MARKED INCREASE IN
CC CANCER SUSCEPTIBILITY. IT IS CHARACTERIZED BY A FAMILIAL
CC PREDISPOSITION TO EARLY ONSET COLORECTAL CARCINOMA (CRC) AND
CC EXTRA-COLONIC CANCERS OF THE GASTROINTESTINAL, UROLOGICAL AND
CC FEMALE REPRODUCTIVE TRACTS. HNPCC IS REPORTED TO BE THE MOST
CC COMMON FORM OF INHERITED COLORECTAL CANCER IN THE WESTERN WORLD.
CC CANCERS IN HNPCC ORIGINATE WITHIN BENIGN NEOPLASTIC POLYPS TIERED
CC ADENOMAS. CLINICALLY, HNPCC IS OFTEN DIVIDED INTO TWO SUBGROUPS.
CC TYPE I: HEREDITARY PREDISPOSITION TO COLORECTAL CANCER, A YOUNG
CC AGE OF ONSET, AND CARCINOMA OBSERVED IN THE PROXIMAL COLON. TYPE
CC II: PATIENTS HAVE AN INCREASED RISK FOR CANCERS IN CERTAIN TISSUES
CC SUCH AS THE UTERUS, OVARY, BREAST, STOMACH, SMALL INTESTINE, SKIN,
CC AND LARYNX IN ADDITION TO THE COLON. DIAGNOSIS OF CLASSICAL HNPCC
CC IS BASED ON THE AMSTERDAM CRITERIA: 3 OR MORE RELATIVES AFFECTED
CC BY COLORECTAL CANCER, ONE A FIRST DEGREE RELATIVE OF THE OTHER
CC TWO; 2 OR MORE GENERATION AFFECTED; 1 OR MORE COLORECTAL CANCERS
CC PRESENTING BEFORE 50 YEARS OF AGE; EXCLUSION OF HEREDITARY
CC POLYPOSIS SYNDROMES. THE TERM "SUSPECTED HNPCC" OR "INCOMPLETE
CC HNPCC" CAN BE USED TO DESCRIBE FAMILIES WHO DO NOT OR ONLY
CC PARTIALLY FULFILL THE AMSTERDAM CRITERIA, BUT IN WHOM A GENETIC
CC BASIS FOR COLON CANCER IS STRONGLY SUSPECTED.
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MTHL/HEXX FAMILY.
CC -1- DATABASE: NAME=hereditary non-polyposis colorectal cancer db;
CC WWW="http://www.nidnt.nl/".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U13695; AA63922.1; -.
DR HSSP: P2367; IBKN.
DR GeneW: HGNC:9121; PMS1.
DR MIM: 600258; -.
DR InterPro: IPR003594; Atphnd_Atpase.
DR InterPro: IPR002099; DNA_mis_repair.
DR InterPro: IPR000910; HMG_12_box.
DR Pfam: PF00505; HMG_box; 1.
DR Pfam: PF01119; DNA_mis_repair; 1.
DR Pfam: PF02518; HATPase_c; 1.
DR SMART: SM00398; HMG; 1.
DR TIGRfam: TIGR00585; mutL; 1.
DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW DNA repair; Anti-oncogene; Nuclear protein;
KW Hereditary nonpolyposis colorectal cancer; Disease mutation.
FT VARIANT 394 394 M -> T (IN INCOMPLETE HNPCC).
FT /FTID=VAR_012967.
FT VARIANT 501 501 G -> R (IN INCOMPLETE HNPCC).
FT /FTID=VAR_012968.
FT SEQUENCE 932 AA; 105830 MW; EC4F402937B61D5F CRC64;

Query Match 3.5%; Score 7; DB 1; Length 932;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 170 VRLSRQL 176
Db 882 VRLSRQL 888

RESULT 10
YA3B_SCHPO STANDARD; PRT; 1294 AA.
ID YA3B_SCHPO
```

AC 009716; 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C18B11.11 in chromosome 1.  
 GN SPAC18B11.11 OR SPAC1F5.01.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NC NCB1\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Soutos J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grympos B.,  
 RA Welfens I., Vansteelt E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wandut R., Purcell B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas R., Rochet M., Gallardin C., Tallada V.A., Garcon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Srpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880(2002).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL: Z50728; CA90595.1; -  
 DR EMBL: Z68136; CA92229.1; -  
 KM Hypothetical protein: Transmembrane.  
 FT TRANSMEM 205 225  
 FT TRANSMEM 277 297 POTENTIAL.  
 FT TRANSMEM 553 573 POTENTIAL.  
 FT TRANSMEM 888 908 POTENTIAL.  
 SQ SEQUENCE 1294 AA: 149192 MW: 82457B940A047469 CRC64;

Query Match 3.5%; Score 7; DB 1; Length 1294;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 TSFSLDF 46  
 DB 460 TSFSLDF 466

RESULT 11  
 PA2\_NAJSF STANDARD; PRT; 16 AA.  
 AC Q10756;

DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase A2 (EC 3.1.1.4) (Phosphatidylcholine 2-acylhydrolase)  
 DE (muscarinic inhibitor) (Fragment).  
 OS Naja sputatrix (Malayan spitting cobra).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 OC Elapidae; Elapinae; Naja.  
 NC NCB1\_TaxID=33626;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Venom;  
 RX MEDLINE=96195757; PubMed=8638927;  
 RA Miyoshi S.-I., Tu A.T.;  
 RT "Phospholipase A2 from Naja sputatrix venom is a muscarinic  
 RT acetylcholine receptor inhibitor."  
 RL Arch. Biochem. Biophys. 328:17-25(1996).  
 CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-  
 CC ACTL GROUPS IN 3-SN-PHOSPHOGLYCERIDES. BLOCKS NEUROMUSCULAR  
 CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE MUSCARINIC  
 CC ACETYLCHOLINE RECEPTOR.  
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-  
 CC acylglycerophosphocholine + a fatty acid anion.  
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.  
 CC HSP; P00598; IPOA.  
 DR InterPro: IPR001211; PhospholipaseA2.  
 DR Pfam: PF00068; Phoslip; 1.  
 DR PROSITE: PS00118; PA2\_HIS; PARTIAL.  
 DR PROSITE: PS00119; PA2\_ASP; PARTIAL.  
 KM Hydrolyase, lipid degradation; Calcium; Venom; Neurotoxin;  
 KM Presynaptic neurotoxin; Postsynaptic neurotoxin.  
 FT NON\_TER 16 16  
 SQ SEQUENCE 16 AA: 1969 MW: 9AC1F9834B585F0 CRC64;

Query Match 3.0%; Score 6; DB 1; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LYQFKN 17  
 DB 2 LYQFKN 7

RESULT 12  
 PA21\_MICNI STANDARD; PRT; 27 AA.  
 AC P21790;  
 ID PA21\_MICNI  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase A2 isozyme 1 (EC 3.1.1.4) (Phosphatidylcholine  
 DE 2-acylhydrolase) (Fragment).  
 OS Micrurus nigrocinctus (Central American coral snake) (Gargantilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 OC Elapidae; Elapinae; Micrurus.  
 NC NCB1\_TaxID=8635;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Venom;  
 RA Mochua-Morales J., Martin B.M., Zamudio F.Z., Possani L.D.;  
 RT "Isolation and characterization of three toxic phospholipases from  
 RT the venom of the coral snake Micrurus nigrocinctus."  
 RL Toxicon 28:616-617(1990).  
 CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-  
 CC ACTL GROUPS IN 3-SN-PHOSPHOGLYCERIDES. INHIBITS NEUROMUSCULAR  
 CC TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE  
 CC TERMINI. ACT PRESYNAPTICALLY.  
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-

```

CC      acylglycerophosphocholine + a fatty acid anion.
CC      -1- COFACTOR: Binds 1 calcium ion per subunit (by similarity).
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
DR      PIR: A35948; A35948.
DR      HSSP: P00598; 1POA.
DR      InterPro: IPR001211; PhospholipaseA2.
DR      Pfam: PF00068; Phoslip; 1.
DR      PROSITE: PS00118; PA2_HIS; PARTIAL.
DR      PROSITE: PS00119; PA2_ASP; PARTIAL.
KW      Hydrolase; Lipid degradation; Calcium; Presynaptic neurotoxin; Venom;
KW      Multigene family.
FT      NON_TER 27.
SQ      SEQUENCE 27 AA; 3314 MW; 38637ECA600F49A0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 LYOFKN 17
DB      2 LYOFKN 7

RESULT 13
PA23_MICNI STANDARD; PRT; 28 AA.
AC      P21792;
DT      01-MAY-1991 (Rel. 18, Last sequence update)
DE      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Phospholipase A2 isozyme 3 (EC 3.1.1.4) (Phosphatidylcholine
DE      2-acylhydrolase) (Fragment).
OS      Micurus nigrocinctus (Central American coral snake) (Gerranillia).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC      Elapidae; Elapinae; Micurus.
RN      NCBI_TaxID=8635;
RX      [1]
RP      SEQUENCE.
RC      TISSUE=Venom;
RA      Mochizuki-Morales J., Martin B.M., Zamudio F.Z., Possani L.D.;
RT      "Isolation and characterization of three toxic phospholipases from
RT      the venom of the coral snake Micurus nigrocinctus.";
RL      Toxicon 28:616-617(1990).
CC      -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-
CC      ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES. INHIBITS NEUROMUSCULAR
CC      TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE
CC      TERMINI. ACT PRESYNAPTICALLY.
CC      -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC      acylglycerophosphocholine + a fatty acid anion.
CC      -1- COFACTOR: Binds 1 calcium ion per subunit (by similarity).
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
DR      PIR: C35948; C35948.
DR      HSSP: P15445; 1A3D.
DR      InterPro: IPR001211; PhospholipaseA2.
DR      Pfam: PF00068; Phoslip; 1.
DR      ProDom: PD000303; PhospholipaseA2; 1.
DR      PROSITE: PS00118; PA2_HIS; PARTIAL.
DR      PROSITE: PS00119; PA2_ASP; PARTIAL.
KW      Hydrolase; Lipid degradation; Calcium; Presynaptic neurotoxin; Venom;
KW      Multigene family.
FT      NON_TER 28.
SQ      SEQUENCE 28 AA; 3394 MW; 315FB012E69098B1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 LYOFKN 17
DB      2 LYOFKN 7

```

```

RESULT 14
LTVB_CHLPN STANDARD; PRT; 97 AA.
ID      LTVB_CHLPN
AC      0928K5;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Late transcription unit B protein.
GN      LTVB OR CPN0333 OR CP0425 OR CPJ0333.
OS      Chlamydia pneumoniae (Chlamydochlamydia pneumoniae).
OC      Bacteria; Chlamydiales; Chlamydiaceae; Chlamydochlamydia.
OX      NCBI_TaxID=83358;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CWL029;
RX      MEDLINE=99206606; PubMed=10192388;
RA      Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA      Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RA      "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RA      Nat. Genet. 21:385-389(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AR39;
RX      MEDLINE=20150255; PubMed=10684935;
RA      Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA      White O., Hickey E.R., Peterson J., Utterback T., Berry K., Bass S.,
RA      Linher K., Weidman J., Khouiri H., Craven B., Bowman C., Dodson R.,
RA      Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA      Eisen J., Fraser C.M.;
RA      "Genome sequences of Chlamydia trachomatis Mofn and Chlamydia
RT      pneumoniae AR39.";
RL      Nucleic Acids Res. 28:1397-1406(2000).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=J138;
RX      MEDLINE=20330349; PubMed=10871362;
RA      Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA      Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT      "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT      from Japan and CWL029 from USA.";
RL      Nucleic Acids Res. 28:2311-2314(2000).
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed, usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: AE001618; AAD18482.1; -
DR      EMBL: AE002203; AAF38268.1; -
DR      EMBL: AP002546; BAA98543.1; -
DR      TIGR: CP0425; -
KW      Complete proteome.
SQ      SEQUENCE 97 AA; 11511 MW; 327AA6511EBF7CA CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 NRRKFL 12
DB      35 NRRKFL 40

RESULT 15
RPOA_BUCAP STANDARD; PRT; 113 AA.
ID      RPOA_BUCAP
AC      069232;
DT      15-DEC-1998 (Rel. 37, Created)

```

```

DT 15-DEC-1998 (Rel. 37, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase
DE alpha chain) (RNA polymerase alpha subunit) (Fragment).
GN RPOA.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxId=98794;
RN [1]
RP SEQUENCE FROM N.A.
RA Munson M.A., Baumann L., Baumann P.;
RT "Buchnera aphidicola, the endosymbiont of aphids, contains genes for
RT four ribosomal RNA proteins, initiation factor-3, and the alpha
RT subunit of RNA polymerase."
RU Curr. Microbiol. 24:23-29(1992).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH
CC IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA' CHAIN AND 1
CC OMEGA CHAIN (BY SIMILARITY).
CC -1- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF
CC CORE RNAP, WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH
CC TRANSCRIPTIONAL REGULATORS.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74510; AAC05400.1; -.
DR HSSP: P00374; 1BDF.
DR InterPro: IPR001700; RNA_pol_A_bac.
DR Pfam: PF01000; RNA_pol_A_bac.1.
DR ProDom: PD001179; RNA_pol_A_bac.1.
KW Transferase; Transcription; DNA-directed RNA polymerase.
FT NON_TER 113 113
SQ SEQUENCE 113 AA; 12415 MW; 033FD3361D5EBF2A CRC64;

```

```

Query Match          3.0%; Score 6; DB 1; Length 113;
Best Local Similarity 100.0%; Pred No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 176 LRRILL 181
   |||||
Db 43 LRRILL 48

```

Search completed: July 7, 2003, 23:43:43  
 Job time : 18 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 23:40:07 ; Search time 20 Seconds  
(without alignments)  
951.731 Million cell updates/sec

Title: US-09-966-880a-8

Perfect score: 198  
Sequence: 1 MSLLMNRKFLYQFKNVW.....ILLPLVEVDLSDAFTLGL 198

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size: 0  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	4.5	1263	2 T13805	spalt-related prot
2	8	4.0	192	2 AD2488	hypothetical prote
3	8	4.0	685	2 T12139	NADH2 dehydrogenas
4	8	4.0	691	2 T12293	NADH2 dehydrogenas
5	8	4.0	699	2 T12648	NADH2 dehydrogenas
6	8	4.0	705	2 T13494	NADH2 dehydrogenas
7	8	4.0	740	1 DENTM5	NADH2 dehydrogenas
8	8	4.0	1178	2 S78475	mannosylphosphoryl
9	7	3.5	90	2 A44977	calmodulin homolog
10	7	3.5	114	2 J05860	polyketide synthas
11	7	3.5	143	2 E70780	hypothetical prote
12	7	3.5	152	2 F98259	probable transcrip
13	7	3.5	152	2 AC3025	transcription regu
14	7	3.5	187	2 F72665	hypothetical prote
15	7	3.5	190	2 AD1418	glutamine amidotra
16	7	3.5	251	2 F83604	hypothetical prote
17	7	3.5	303	2 S75557	hypothetical prote
18	7	3.5	350	2 D71817	cytochrome-c perox
19	7	3.5	350	2 E64702	cytochrome-c perox
20	7	3.5	353	2 C96573	protein F12M16.19
21	7	3.5	359	2 D88940	protein C05E4.1 [1
22	7	3.5	368	2 E69796	two-component resp
23	7	3.5	375	2 AB3165	conserved hypothet
24	7	3.5	481	2 B75480	hypothetical prote
25	7	3.5	500	2 B81250	probable site-spec
26	7	3.5	521	2 S34184	cytochrome P450 1A
27	7	3.5	522	2 B75062	flagella accessory
28	7	3.5	615	2 S77084	ABC-type transport
29	7	3.5	889	2 H64506	probable retroelem

30	7	3.5	932	2 S47597	mult. protein homol
31	7	3.5	1294	2 S6307	hypothetical 149.2
32	6	3.0	16	2 S65520	phospholipase A2 (
33	6	3.0	27	2 A35948	phospholipase A2 (
34	6	3.0	28	2 C35948	phospholipase A2 (
35	6	3.0	31	2 J00008	cytotoxin Xlri - b
36	6	3.0	67	2 C96765	hypothetical prote
37	6	3.0	87	2 G86631	hypothetical prote
38	6	3.0	88	2 T24638	hypothetical prote
39	6	3.0	91	2 E71938	hypothetical prote
40	6	3.0	94	2 G84269	30S ribosomal prot
41	6	3.0	97	2 E86532	Ltub protein (limp
42	6	3.0	97	2 C72090	conserved hypothet
43	6	3.0	100	2 A97833	hypothetical prote
44	6	3.0	101	2 AB1916	hypothetical prote
45	6	3.0	101	2 AC3161	DNA-invertase [amp

ALIGNMENTS

RESULT 1  
T13805  
spalt-related protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13805  
R:Barrio, R.; Shea, M.J.; Canull, J.; Lipkow, K.; Gaul, U.; Frommer, G.; Schuh, R.;  
submitted to the EMBL Data Library, August 1996  
A:Description: The spalt-related gene of Drosophila melanogaster is a member of an  
A:Reference number: 217768  
A:Accession: T13805  
A:status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1263 <BAR>  
A:Cross-references: EMBL:X07653; NID:e1004366; PID:e267584; PIDN:CA68937.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0000287  
A:introns: 862/1; 915/3; 1003/1; 1080/3; 1146/3

Query Match 4.5%; Score 9; DB 2; Length 1263;  
Best Local Similarity 100.0%; Pred. No. 0.58;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	117	EDRKAREPG 125
DB	237	EDRKAREPG 245

RESULT 2  
AD2488  
hypothetical protein al17084 [imported] - Nostoc sp. (strain PCC 7120) plasmid pcc712  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AD2488  
R:Kanezo, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata  
DNA Res. 9, 205-213, 2001  
A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AD2488  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <KUR>  
A:Cross-references: GB:BA000020; PIDN:BA878168.1; PID:gt135622; GSPDB:GN00180  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: al17084  
A:Genome: plasmid  
Query Match 4.0%; Score 8; DB 2; Length 192;  
Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 VDDLRDAF 193  
|||||

Db 104 VDDLRDAF 111

#### RESULT 3

T12139 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Alsobia sp. chloroplast (fragment)  
C:Species: chloroplast Alsobia sp.

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 03-Jun-2002

C:Accession: T12139

R:Smith, J.F.; Kresge, M.E.; Moller, M.; Cronk, Q.C.

submitted to the EMBL Data Library, July 1997

A:Reference number: 217428

A:Accession: T12139

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-685 <SMI>

A:Cross-references: EMBL:AF013689; NID:g2431875; PID:g2431876; PIDN:AA871143.1

C:Genetics:

A:Genome: chloroplast

A:Note: ndhf

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C:Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 4.0%; Score 8; DB 2; Length 685;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### RESULT 4

T12293 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Bellonia sp. 'Evans s.n. chloroplast'  
C:Species: chloroplast Bellonia sp. 'Evans s.n.

C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 03-Jun-2002

C:Accession: T12293

R:Smith, J.F.; Atkinson, S.

submitted to the EMBL Data Library, December 1997

A:Description: Phylogenetic analysis of the tribes Gloxiniaceae and Gesneriaceae (Gesneriaceae)

A:Reference number: 217484

A:Accession: T12293

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-691 <SMI>

A:Cross-references: EMBL:AF040144; NID:g2791782; PID:g2791783; PIDN:AA86929.1

C:Genetics:

A:Genome: chloroplast

A:Note: ndhf

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C:Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 4.0%; Score 8; DB 2; Length 691;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
|||||

Db 72 FSLDFGYL 79

RESULT 5  
T12648 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Brillantaisia lanium chloroplast  
C:Species: chloroplast Brillantaisia lanium  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 03-Jun-2002  
C:Accession: T12648

R:Scotland, R.W.; Sweere, J.A.; Reeves, P.A.; Olmstead, R.G.

submitted to the EMBL Data Library, July 1994

A:Description: Higher level systematics of Acanthaceae determined by chloroplast DNA

A:Reference number: 217564

A:Accession: T12648

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-699 <SCO>

A:Cross-references: EMBL:U12654; NID:g607918; PID:g639914; PIDN:AA61718.1

C:Genetics:

A:Genome: chloroplast

A:Note: ndhf

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C:Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 4.0%; Score 8; DB 2; Length 699;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### RESULT 6

T13494 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Nematanthus hirsutus chloroplast  
C:Species: chloroplast Nematanthus hirsutus

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002

C:Accession: T13494

R:Olmstead, R.G.; Reeves, P.A.

Ann. Mo. Bot. Gard. 82, 176-193, 1995

A:Title: Evidence for the polyphyly of the Scrophulariaceae based on chloroplast rbcL

A:Reference number: 217559

A:Accession: T13494

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-705 <OLM>

A:Cross-references: EMBL:L36404; NID:g703198; PID:g703199; PIDN:AA84497.1

C:Genetics:

A:Genome: chloroplast

A:Note: ndhf

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C:Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 4.0%; Score 8; DB 2; Length 705;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### RESULT 7

DENTIN5

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - common tobacco chloroplast

N:Alternate names: NADH-ubiquinone oxidoreductase chain 5; ndh5 protein

C:Species: chloroplast Nicotiana glauca (common tobacco)

C:Date: 30-Jun-1987 #sequence\_revision 14-Aug-1998 #text\_change 03-Jun-2002

C:Accession: S37352; A00454

R:Olmstead, R.G.; Sweere, J.A.; Wolfe, K.H.

Plant Mol. Biol. 22, 1191-1193, 1993

A:Title: Ninety extra nucleotide in ndhF gene of tobacco chloroplast DNA: a summary o

A:Reference number: S37352; M01D:94003079; PMID:8400137

A:Accession: S37352

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-740 <OLM>

A:Cross-references: GB:U14953; NID:g295333; PIDN:AA84685.1; PID:g295334

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993

R:Sugawara, M.  
submitted to the EMBL Data Library, August 1986

A:Reference number: A00149  
A:Accession: A00454  
A:Molecule type: DNA  
A:Residues: 1-578,609-691,'TD',694-740 <SUG>  
A:Experimental source: cv. Bright Yellow 4  
R:Shinozaki, K.; Ohme, M.; Tanaka, M.; Wakasugi, T.; Hayashida, N.; Matsubayashi, T.; Zeng, H.; Kamogashira, T.; Yamada, K.; Kusuda, J.; Takaiwa, F.; Kato, A.; Tohdo, N.; Shimizu, J., 2043-2049, 1986  
A:Title: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization, gene organization, sites, features  
A:Contents: annotation; gene organization, sites, features  
C:Genetics:  
A:Gene: ndhF  
A:Genome: chloroplast  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
C:Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 4.0%; Score 8; DB 1; Length 740;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
|||||||  
DB 81 FSLDFGYL 88

RESULT 8  
S78475  
manosylphosphorylation protein MN4 - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein YKL200C; protein YKL201C  
C:Species: *Saccharomyces cerevisiae*  
C>Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 19-Apr-2002  
C:Accession: S78475; S38037; S38038  
R:Odani, T.; Shimizu, Y.; Yoshitani, J.  
Submitted to the EMBL Data Library, January 1996  
A:Description: Cloning and Analysis of the MN4 Gene Required for Phosphorylation of N-1  
A:Reference number: S78475  
A:Accession: S78475  
A:Molecule type: DNA  
A:Residues: 1-1178 <DNA>  
A:Cross-references: EMBL:D83006; NID:g1752735; PID:d1012343; PID:g1752736; MIPS:YKL201C  
A:Note: this is a revision to the sequence from reference S38024  
R:Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.; R  
Submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38024  
A:Accession: S38037  
A:Molecule type: DNA  
A:Residues: 121-249,'ATGLOT',255,'MALLRD',262-390,'LRISSEN',398-515,'LG' <MAI>  
A:Cross-references: EMBL:D28200; NID:g486357; PID:g486359; MIPS:YKL201C  
A:Experimental source: strain S288C  
A:Note: this sequence has been revised in reference S78475  
A:Note: this sequence has been revised in reference S78475  
A:Accession: S38038  
A:Molecule type: DNA  
A:Residues: 640-1178 <MAN>  
A:Cross-references: EMBL:D28201; NID:g486357; PID:g486359; MIPS:YKL201C  
A:Experimental source: strain S288C  
A:Note: this sequence has been revised in reference S78475  
A:Note: this sequence has been revised in reference S78475  
C:Genetics:  
A:Gene: SGD:MNN4; MNN4  
A:Cross-references: MIPS:YKL201C; SGD:S0001684  
A:Map position: 11L  
C:Function:  
A:Description: required for phosphorylation of N-linked oligosaccharides  
C:Keywords: transmembrane protein  
P:28-44/Domain: transmembrane #status predicted <TMM>

Query Match 4.0%; Score 8; DB 2; Length 1178;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 LRRLLPL 183

DB 25 LRRLLPL 32  
|||||||

RESULT 9  
A44977  
calmodulin homolog - fluke (*Schistosoma mansoni*) (fragment)  
C:Species: *Schistosoma mansoni*  
C>Date: 28-Apr-1993 #sequence\_revision 28-Apr-1993 #text\_change 13-Aug-1999  
C:Accession: A44977  
R:Havecroft, J.C.; Huggins, M.C.; Dunne, D.W.; Taylor, D.W.  
Mol. Biochem. Parasitol. 38, 211-220, 1990  
A:Title: Characterisation of Sm20, a 20-kilodalton calcium-binding protein of *Schisto*  
A:Reference number: A44977; MUID:90220753; PMID:2335706  
A:Accession: A44977  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-90 <HAV>  
A:Cross-references: GS:M28888; NID:g161095; PID:AAA2926.1; PID:g161096  
C:Superfamily: calmodulin; calmodulin repeat homology  
C:Keywords: EF hand  
P:58-90/Domain: calmodulin repeat homology <EFA>

Query Match 3.5%; Score 7; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 DLRFDAFR 194  
|||||||  
DB 26 DLRFDAFR 32

RESULT 10  
JC5860  
polyketide synthase (EC 2.3.1.-) chain 11 - *Actinobadura hibisca*  
C:Species: *Actinobadura hibisca*  
C>Date: 05-Mar-1998 #sequence\_revision 13-Mar-1998 #text\_change 20-Jun-2000  
C:Accession: JC5860  
R:Dairi, T.; Hamano, Y.; Igarashi, Y.; Furumai, T.; Oki, T.  
Bioosci. Biotechnol. Biochem. 61, 1445-1453, 1997  
A:Title: Cloning and nucleotide sequence of the putative polyketide synthase genes for  
A:Reference number: JC5860; MUID:97480928; PMID:9339544  
A:Accession: JC5860  
A:Molecule type: DNA  
A:Residues: 1-114 <DAI>  
A:Cross-references: DDBJ:D87924; NID:g2580441; PID:BA23154.1; PID:g2580452  
A:Comment: This enzyme catalyzes repeated condensation cycles of acyl-CoA, affecting  
C:Genetics:  
A:Gene: pms11  
C:Superfamily: tetracycline-biosynthetic D-ring cyclase  
C:Keywords: transferase

Query Match 3.5%; Score 7; DB 2; Length 114;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 VRLSROL 176  
|||||||  
DB 73 VRLSROL 79

RESULT 11  
E70780  
hypothetical protein RV0880 - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 28-Jul-2000  
C:Accession: E70780  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: deciphering the biology of *Mycobacterium tuberculosis* from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: E70780  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-143 <COL>  
A:Cross-references: GB:J273101; GB:AL123456; NID:g3261565; PIDN:CAA97387.1; PID:g1314019  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV0880  
C:Superfamily: Mycobacterium leprae hypothetical protein MCB57.31

Query Match 3.5%; Score 7; DB 2; Length 143;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 171 RLSROLR 177  
18 RLSROLR 24

RESULT 12  
F98259  
probable transcription regulator PA4508 [imported] - Agrobacterium tumefaciens (strain C  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: F98259  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: F98259  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-152 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK89600.1; PID:g15159491; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_2051  
A:Map position: linear chromosome  
C:Superfamily: regulatory protein asnc

Query Match 3.5%; Score 7; DB 2; Length 152;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 SVRLSRQ 175  
73 SVRLSRQ 79

RESULT 13  
AC3025  
transcription regulator, AsnC family lrp [imported] - Agrobacterium tumefaciens (strain  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AC3025  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Giller, W.; Grant, C.; Genthner, D.; Kulyavin, T.; Levy, R.; Li, M.; McClell  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AC3025  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-152 <KUR>  
A:Cross-references: GB:AE008669; PIDN:AA144617.1; PID:g17742239; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: lrp  
A:Map position: linear chromosome

C:Superfamily: regulatory protein asnc

Query Match 3.5%; Score 7; DB 2; Length 152;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 SVRLSRQ 175  
73 SVRLSRQ 79

RESULT 14  
F72665  
hypothetical protein APE0749 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: F72665  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: F72665  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-187 <KAW>  
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79726.1; PID:d1043512; PID:g  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0749

Query Match 3.5%; Score 7; DB 2; Length 187;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 124 EGLRLH 130  
125 EGLRLH 131

RESULT 15  
AD1418  
glutamine amidotransferase homolog lmo2749 [imported] - Listeria monocytogenes (strai  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AD1418  
R:Glaser, P.; Frangouli, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entlian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madeno, E.; Maitournam, A.;  
Ok, C.; Schueller, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1418  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-190 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00962.1; PID:g16412249; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2749  
C:Superfamily: glutamine amidotransferase; trpG homology

Query Match 3.5%; Score 7; DB 2; Length 190;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 SATSFSL 44  
30 SATSFSL 36

Mon Jul 14 14:47:11 2003

us-09-966-880a-8.olg.rpr

Page 5

Search completed: July 7, 2003, 23:45:01  
job time : 22 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 23:39:02 ; Search time 31 Seconds

(without alignments)  
1316.043 Million cell updates/sec

Title: us-09-966-880a-8

Perfect score: 198

Sequence: 1 MSLLMNRKFLYQFNVRW.....ILLPLVEVDLDRAPRTIGL 198

Scoring table: OLIGO

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SPTREMBL\_21.\*

1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_virus:\*  
16: sp.\_bacteria:\*  
17: sp.\_archae:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	198	100.0	198	4	Q9GZX7
2	42	21.2	198	11	Q9WVE0
3	9	4.5	1263	5	P91639
4	9	4.5	1267	5	Q9VKH3
5	8	4.0	192	16	Q8YL51
6	8	4.0	291	2	Q85677
7	8	4.0	410	8	Q9MS87
8	8	4.0	412	8	Q98472
9	8	4.0	417	8	Q95683
10	8	4.0	420	16	Q9RL01
11	8	4.0	521	8	Q9IKN8
12	8	4.0	646	8	Q32382
13	8	4.0	663	8	Q9MS52
14	8	4.0	675	8	Q32213
15	8	4.0	684	8	Q9MUG6
16	8	4.0	685	8	Q19822

17	8	4.0	690	8	Q9MUB1	Q9mub1 solanum mac
18	8	4.0	691	8	Q47006	Q47006 belionia sp
19	8	4.0	693	8	Q95669	Q95669 solanum abu
20	8	4.0	694	8	Q95680	Q95680 solanum lac
21	8	4.0	694	8	Q9MUA6	Q9mua6 solanum jam
22	8	4.0	694	8	P92308	P92308 lycianthes
23	8	4.0	695	8	Q32158	Q32158 datura stra
24	8	4.0	695	8	Q32476	Q32476 juanilloa a
25	8	4.0	695	8	Q32508	Q32508 lycium cest
26	8	4.0	695	8	Q32602	Q32602 mandragora
27	8	4.0	695	8	Q32654	Q32654 nicotiana a
28	8	4.0	695	8	Q32687	Q32687 nicotiana ph
29	8	4.0	695	8	Q32700	Q32700 nicotiana spat
30	8	4.0	695	8	Q32779	Q32779 physalis al
31	8	4.0	695	8	Q33102	Q33102 solandra gr
32	8	4.0	695	8	Q95671	Q95671 solanum arb
33	8	4.0	695	8	Q95672	Q95672 solanum avi
34	8	4.0	695	8	Q95684	Q95684 solanum pse
35	8	4.0	695	8	Q95682	Q95682 solanum ply
36	8	4.0	695	8	Q95685	Q95685 solanum ros
37	8	4.0	695	8	Q95686	Q95686 solanum sea
38	8	4.0	695	8	Q95689	Q95689 solanum wal
39	8	4.0	695	8	Q95674	Q95674 solanum bet
40	8	4.0	695	8	Q95659	Q95659 jalomata p
41	8	4.0	695	8	Q9MUB8	Q9mub8 solanum adh
42	8	4.0	695	8	Q9MUB7	Q9mub7 solanum app
43	8	4.0	695	8	Q9MUB6	Q9mub6 normania tr
44	8	4.0	695	8	Q9MUB5	Q9mub5 solanum pal
45	8	4.0	695	8	Q9MUB4	Q9mub4 triguera os

## ALIGNMENTS

RESULT 1

Q9GZX7 PRELIMINARY; PRI; 198 AA.

AC Q9GZX7; 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Activation-induced cytidine deaminase.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

ON [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-20408890; PubMed-10950930;

RA Muto T., Kuramatsu M., Taniwaki M., Kinoshita K., Honjo T.,

RT "Isolation, tissue distribution and chromosomal localization of the

RL human activation-induced cytidine deaminase (Aid) gene.";

RN genomics 68:85-88(2000).

[2]

RP SEQUENCE FROM N.A.

RX MEDLINE-20460541; PubMed-11007475;

RA Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O.,

RA Catalan N., Forveille M., Dufourcq-Lageouse R., Gennery A.,

RA Tezcan I., Ersoy F., Kayserili H., Ugazio A.G., Brousse N.,

RA Muramatsu M., Notarangelo L.D., Kinoshita K., Honjo T., Fischer A.,

RA Durandy A.,

RT "Activation-induced cytidine deaminase (Aid) deficiency causes the

RL autosomal recessive form of the Hyper-IgM syndrome (HIGM2).";

RN Cell 102:565-575(2000).

[3]

RP SEQUENCE FROM N.A.

RC TISSUE-B-CELL;

RA Strausberg R.;

RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL, AB040431; BAB12721.1; -

DR EMBL, AB040430; BAB12720.1; -

DR EMBL, BC006296; AAH06296.1; -

DR InterPro: IPR002125; dCMP/cyt\_deam.  
 DR PROSITE: PS00903; CYT\_DCMP\_DEAMINASES; UNKNOWN\_1.  
 SO SEQUENCE 198 AA; 23953 MW; 3C27BB143DB184A9 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 198; DB 4; Length 198;  
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLNRRKFLYQFNVRNAKRRRETYLCYVKRRDSATSFSLDFGYLRNKGCHVELL 60  
 DB 1 MSLLNRRKFLYQFNVRNAKRRRETYLCYVKRRDSATSFSLDFGYLRNKGCHVELL 60

QY 61 FLRTISWDLPDRCRYVWFTSMSPCYDCARVAPFLKGNPULSRITARLYFCEDRK 120  
 DB 61 FLRTISWDLPDRCRYVWFTSMSPCYDCARVAPFLKGNPULSRITARLYFCEDRK 120

QY 121 AEPGLRLHRAAGVQVQIMTFKDYFCWNTFVNHRTFAKAGLHENSVRSLROLRIL 180  
 DB 121 AEPGLRLHRAAGVQVQIMTFKDYFCWNTFVNHRTFAKAGLHENSVRSLROLRIL 180

QY 181 LPLEYVDLDRDAFRTIGL 198  
 DB 181 LPLEYVDLDRDAFRTIGL 198

RESULT 2  
 Q9WVE0 PRELIMINARY; PRT; 198 AA.  
 AC Q9WVE0;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Activation-induced cytidine deaminase.  
 GN AICDA OR AID.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99303612; PubMed=10373455;  
 RA Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K.,  
 RA Davidson N.O., Honjo T.;  
 RT "Specific expression of activation-induced cytidine deaminase (AID), a  
 RT novel member of the RNA-editing deaminase family in germinal center B  
 RT cells.";  
 RT J. Biol. Chem. 274:18470-18476(1999).  
 DR EMBL: AF132979; AAD41793.1; -.  
 DR MGD: MGI:1342279; Aicda.  
 DR InterPro: IPR002125; dCMP/cyt\_deam.  
 DR PROSITE: PS00903; CYT\_DCMP\_DEAMINASES; UNKNOWN\_1.  
 SO SEQUENCE 198 AA; 24030 MW; 18A3BA10CA54BB2 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 42; DB 11; Length 198;  
 Matches 42; Conservative 0; Mismatches 56-37; Indels 0; Gaps 0;

QY 54 GCHVELLFLRYISWDLPDRCRYVWFTSMSPCYDCARVAV 95  
 DB 54 GCHVELLFLRYISWDLPDRCRYVWFTSMSPCYDCARVAV 95

RESULT 3  
 P91639 PRELIMINARY; PRT; 1263 AA.  
 AC P91639;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Zinc-finger protein SALR (SPALT-related protein).  
 GN SALR OR CG4881.  
 OS Drosophila; melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RA Barrio R., Shea M.J., Carrull J., Lipkow K., Gaul U., Frommer G.,  
 RA Schu R., Jackle H., Katatos F.C.;  
 RT "The spalt-related gene of Drosophila melanogaster is a member of an  
 RT ancient gene family, defined by the adjacent, region-specific homeotic  
 RT gene spalt.";  
 RT Dev. Genes Evol. 206:315-325(1996).  
 CC -; SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: Y07653; CAA68937.1; -.  
 DR HSSP: P15822; 1BBO.  
 DR Flybase: FBgn0000287; salr.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR008222; ZnF\_C2H2.  
 DR Pfam: PF00096; ZF-C2H2; 8.  
 DR PRINTS: PR00048; ZINC\_FINGER.  
 DR SMART: SM00355; ZnF\_C2H2; 8.  
 DR PROSITE: PS00290; IG\_MHC; UNKNOWN\_1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE: PS01577; ZINC\_FINGER\_C2H2\_2; 7.  
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.  
 SO SEQUENCE 1263 AA; 138991 MW; 2025B3BE6741622 CRC64;

Query Match  
 Best Local Similarity 4.5%; Score 9; DB 5; Length 1263;  
 Matches 9; Conservative 0; Mismatches 1.3; Indels 0; Gaps 0;

QY 117 EDKRAPEEG 125  
 DB 237 EDKRAPEEG 245

RESULT 4  
 Q9VKH3 PRELIMINARY; PRT; 1267 AA.  
 AC Q9VKH3;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE SALR protein.  
 GN SALR OR CG4881.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chape M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintinas S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarik C., Morris J., Mosterfi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Stykstra R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2165-2195(2000).  
 DR EMBL; AF003632; AAF53096.1; -.  
 DR HSSP; P15822; IBB0.  
 DR FlyBase; FBgn0000287; salr.  
 DR InterPro; IPR003006; Iq\_MHC.  
 DR InterPro; IPR000822; znf\_C2H2.  
 DR Pfam; PF00096; znf\_C2H2; 8.  
 DR SMART; SM00355; znf\_C2H2; 8.  
 DR PROSITE; PS00290; Iq\_MHC; UNKNOWN\_1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 7.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 SQ SEQUENCE 1267 AA; 139540 MW; 4510DB1C0F47829 CRC64;

Query Match 4.5%; Score 9; DB 5; Length 1267;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 EDRKAPEPG 125  
 Db 237 EDRKAPEPG 245

## RESULT 5

Q8YL51 PRELIMINARY; PRT; 192 AA.  
 AC Q8YL51;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein A117084.  
 GN A117084.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Plasmid PCC7120alpha.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kunitz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium *Anabaena* sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003600; BAB78168.1; -.  
 DR InterPro; IPR001601; Methyltransf.  
 KW Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 192 AA; 21971 MW; 8BBA3178B34BFC CRC64;

Query Match 4.0%; Score 8; DB 16; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 186 VDDLRFAD 193  
 Db 104 VDDLRFAD 111

## RESULT 6

OC 085677 PRELIMINARY; PRT; 291 AA.  
 ID 085677;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Bvgr.  
 GN Bvgr.  
 OS *Bordetella pertussis*.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCBI\_TaxID=520;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TOHAMA I;  
 RX MEDLINE=98196709; PubMed=9537363;  
 RA Merkel T.J., Barros C., Stibitz S.;  
 RT "Characterization of the bvgR locus of *Bordetella pertussis*.";  
 RL J. Bacteriol. 180:1682-1690(1998).  
 DR EMBL; AF071567; AAC23902.1; -.  
 DR InterPro; IPR001633; EAL.  
 DR Pfam; PF00563; EAL; 1.  
 DR SMART; SM00052; DUF2; 1.  
 SQ SEQUENCE 291 AA; 32102 MW; E392CB5FC70DB22 CRC64;

Query Match 4.0%; Score 8; DB 2; Length 291;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 LRLHRAG 133  
 Db 174 LRLHRAG 181

## RESULT 7

Q9MS87 PRELIMINARY; PRT; 410 AA.  
 AC Q9MS87;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE NADH dehydrogenase subunit F (Fragment).  
 GN NDHF.  
 OS *Hydrastis canadensis*.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
 OC Ranunculaceae; Hydrastis.  
 OX NCBI\_TaxID=13569;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Graham S.W., Reeves P.A., Burns A., Olmstead R.G.;  
 RT "Long branches in the seed plants and the root of the angiosperms.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
 DR EMBL; AF238055; AAF90041.2; -.  
 DR InterPro; IPR001750; Oxidored\_q1.  
 DR InterPro; IPR001516; Oxidored\_q1\_N.  
 DR Pfam; PF00361; Oxidored\_q1; 1.  
 DR Pfam; PF00662; Oxidored\_q1\_N; 1.  
 KW Chloroplast; NAD; Oxidoreductase; Plastocyanine.  
 FT NON\_TER 1 410  
 SQ SEQUENCE 410 AA; 45353 MW; FMA3624BA4155C472 CRC64;

Query Match 4.0%; Score 8; DB 8; Length 410;



Best Local Similarity 100.0%; Pred. No. 5.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
|||||  
Db 52 FSLDFGYL 59

## RESULT 8

098472 ID 098472 PRELIMINARY; PRT; 412 AA.

AC 098472; 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NADH dehydrogenase subunit F (Fragment).

GN NDHF.  
OS Aucuba japonica.

OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Garryales; Aucubaceae; Aucuba.

NCBI\_TaxID=16901;  
RX [1]

RP SEQUENCE FROM N.A.  
RA Roels P., Smets E.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

CC -1 CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
DR EMBL; AF060158; AAC64348.1; -;  
DR InterPro: IPR001750; Oxidored\_g1.

DR InterPro: IPR001516; Oxidored\_g1\_N.  
DR Pfam: PF00361; Oxidored\_g1; 1.  
DR Pfam: PF00662; Oxidored\_g1\_N; 1.  
KW Chloroplast; NAD; Oxidoreductase; Plastoquinone.

FT NON\_TER 1  
FT 412 412  
SQ SEQUENCE 412 AA; 45616 MW; C0A3251174071738 CRC64;

Query Match 4.0%; Score 8; DB 8; Length 412;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
|||||  
Db 53 FSLDFGYL 60

## RESULT 9

095683 ID 095683 PRELIMINARY; PRT; 417 AA.

AC 095683; 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NADH dehydrogenase subunit (Fragment).

GN NDHF.  
OS Solanum physallifolium.

OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

NCBI\_TaxID=45836;  
RX [1]

RP SEQUENCE FROM N.A.  
RA Bohs L., Olmstead R.;  
RL "Solanum phylogeny inferred from chloroplast DNA sequence data.";  
RT Fourth Int. Solanaceae Conf. Proc. 0:0-0(0).

CC -1 CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
DR EMBL; U47421; AAB09284.1; -;  
DR InterPro: IPR001750; Oxidored\_g1.  
DR InterPro: IPR001516; Oxidored\_g1\_N.  
DR Pfam: PF00361; Oxidored\_g1; 1.  
DR Pfam: PF00662; Oxidored\_g1\_N; 1.

KW Chloroplast; NAD; Oxidoreductase; Plastoquinone.  
FT NON\_TER 1  
SQ SEQUENCE 417 AA; 46556 MW; 32CA895132354DA CRC64;

Query Match 4.0%; Score 8; DB 8; Length 417;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
|||||  
Db 73 FSLDFGYL 80

## RESULT 10

09RL01 ID 09RL01 PRELIMINARY; PRT; 420 AA.

AC 09RL01; 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative transmembrane transport protein.

GN SC00317 OR SC559.26C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomyces.

NCBI\_TaxID=1902;  
RX [1]

RP SEQUENCE FROM N.A.  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RC STRAIN=A3(2) / M145;

RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Candara G., Chen C.W., Collins M.,  
RA Croftin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;

"Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).

CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; AL117385; CAB55674.1; -;  
KW Transmembrane.

SQ SEQUENCE 420 AA; 42250 MW; 87D5EB267AFD8A8D CRC64;

Query Match 4.0%; Score 8; DB 16; Length 420;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 LRRLLPL 183  
|||||  
Db 214 LRRLLPL 221

## RESULT 11

09TKN8 ID 09TKN8 PRELIMINARY; PRT; 521 AA.

AC 09TKN8; 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE NADH dehydrogenase subunit (Fragment).

GN NDHF.  
OS Exacum affine.

OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Gentianales; Gentianaceae; Exacum.

NCBI\_TaxID=13525;  
RX [1]

RP SEQUENCE FROM N.A.  
RA MEDLINE=20340987; PubMed=10877943;

RA Olmstead R.G., Kim K.J., Jansen R.K., Magstaff S.J.:  
 "The phylogeny of the asteridae sensu lato based on chloroplast ndhF  
 gene sequences.";  
 RL Mol. Phylogenet. Evol. 16:96-112(2000).  
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
 DR EMBL: AF147710; AAF03769.1; -  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR002128; Oxidored\_q1.C.  
 DR InterPro: IPR001516; Oxidored\_q1.N.  
 DR Pfam: PF00361; oxidored\_q1; 1.  
 DR Pfam: PF01010; oxidored\_q1.C; 1.  
 DR Pfam: PF00662; oxidored\_q1.N; 1.  
 KW Chloroplast; NAD; Oxidoreductase; Plastocyanine.  
 FT NON\_TER 1  
 FT SEQUENCE 521 AA; 58897 MW; 89721CB90B41B1CB CRC64;  
 SO  
 Query Match 4.0%; Score 8; DB 8; Length 521;  
 Best Local Similarity 100.0%; Pred. No. 7.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
 |||||  
 Db 73 FSLDFGYL 80

RESULT 12  
 Q32382 PRELIMINARY; PRT; 646 AA.  
 AC Q32382;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE NADH dehydrogenase subunit (Fragment).  
 GN NDMF.  
 OS Hemigraphis alternata.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Acanthaceae; Hemigraphis.  
 CX NCBI\_TaxID=34269;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Scotland R.W., Sweeney J.A., Reeves P.A., Olmstead R.G.;  
 RT "Higher level systematics of Acanthaceae determined by chloroplast DNA  
 sequences.";  
 RL Am. J. Bot. 0:0-0(1995).  
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
 DR EMBL: U12660; AAG61724.1; -  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR002128; Oxidored\_q1.C.  
 DR InterPro: IPR001516; Oxidored\_q1.N.  
 DR Pfam: PF00361; oxidored\_q1; 1.  
 DR Pfam: PF01010; oxidored\_q1.C; 1.  
 DR Pfam: PF00662; oxidored\_q1.N; 1.  
 KW Chloroplast; NAD; Oxidoreductase; Plastocyanine.  
 FT NON\_TER 1  
 FT SEQUENCE 646 AA; 73531 MW; 539878F8B4C67919 CRC64;  
 SO  
 Query Match 4.0%; Score 8; DB 8; Length 646;  
 Best Local Similarity 100.0%; Pred. No. 8.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
 |||||  
 Db 23 FSLDFGYL 30

RESULT 13  
 Q9MS52 PRELIMINARY; PRT; 663 AA.  
 AC Q9MS52;  
 Q9MS52;

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE NADH dehydrogenase (Fragment).  
 GN NDMF.  
 OS Goyazia rupicola.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Gesneriaceae; Goyazia.  
 CX NCBI\_TaxID=125973;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith J.F.;  
 RT "The phylogenetic relationships of Lemnocarpos and Goyazia  
 (Gesneriaceae) based on ndhF sequences.";  
 RL Ann. Mo. Bot. Gard. 0:0-0(2000).  
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
 DR EMBL: AF257485; AAF75262.1; -  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR002128; Oxidored\_q1.C.  
 DR InterPro: IPR001516; Oxidored\_q1.N.  
 DR Pfam: PF00361; oxidored\_q1; 1.  
 DR Pfam: PF01010; oxidored\_q1.C; 1.  
 DR Pfam: PF00662; oxidored\_q1.N; 1.  
 KW Chloroplast; NAD; Oxidoreductase; Plastocyanine.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT SEQUENCE 663 AA; 74858 MW; BB002498FACE322AE CRC64;  
 SO  
 Query Match 4.0%; Score 8; DB 8; Length 663;  
 Best Local Similarity 100.0%; Pred. No. 8.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
 |||||  
 Db 74 FSLDFGYL 81

RESULT 14  
 Q32213 PRELIMINARY; PRT; 675 AA.  
 AC Q32213;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE NADH dehydrogenase subunit (Fragment).  
 GN NDMF.  
 OS Eremomastax speciosa.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Acanthaceae; Eremomastax.  
 CX NCBI\_TaxID=37794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Scotland R.W., Sweeney J.A., Reeves P.A., Olmstead R.G.;  
 RT "Higher level systematics of Acanthaceae determined by chloroplast DNA  
 sequences.";  
 RL Am. J. Bot. 0:0-0(1995).  
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
 DR EMBL: U12659; AAG61723.1; -  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR002128; Oxidored\_q1.C.  
 DR InterPro: IPR001516; Oxidored\_q1.N.  
 DR Pfam: PF00361; oxidored\_q1; 1.  
 DR Pfam: PF01010; oxidored\_q1.C; 1.  
 DR Pfam: PF00662; oxidored\_q1.N; 1.  
 KW Chloroplast; NAD; Oxidoreductase; Plastocyanine.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT SEQUENCE 675 AA; 76873 MW; 943D9622D98FA693 CRC64;  
 SO

Query Match 4.0%; Score 8; DB 8; Length 675;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
 |||||  
 DB 50 FSLDFGYL 57

## RESULT 15

Q9MUG6 PRELIMINARY; PRT; 684 AA.  
 AC Q9MUG6;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE NADH dehydrogenase subunit F (Fragment).  
 GN NDHF.  
 OS Nematanthus albus.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Gesneriaceae; Nematanthus.  
 OX NCBI\_TaxID=122672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith J.F.;  
 RT "Phylogenetic resolution within the tribe Episcleae (Gesneriaceae):  
 RT congruence of ITS and ndhF sequences from parsimony and maximum-  
 RT likelihood analyses."  
 RL Am. J. Bot. 0:0-0(2000).  
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
 DR EMBL: AF206197; AAF70600.1;  
 DR InterPro: IPR001750; Oxidored\_g1.  
 DR InterPro: IPR002128; Oxidored\_g1\_C.  
 DR Pfam: PF00361; Oxidored\_g1; 1.  
 DR Pfam: PF01010; Oxidored\_g1\_C; 1.  
 DR Pfam: PF00662; Oxidored\_g1\_N; 1.  
 DR Chloroplast; NAD; Oxidoreductase; Plastocyanine.  
 KW NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 684 AA; 7753 MW; 4DBB6FEA67DE6C1B CRC64;  
 Query Match 4.0%; Score 8; DB 8; Length 684;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FSLDFGYL 49  
 |||||  
 DB 74 FSLDFGYL 81

Search completed: July 7, 2003, 23:44:27  
 Job time : 32 secs



KW	major histocompatibility class II deficiency disease;
RN	auto immunodeficiency syndrome; IgG subclass selection disorder; ss.
XX	
OS	Homo sapiens.
XX	
FH	Location/Qualifiers
FT	80..676
CD5	/tag= a
FT	/product= "activation-induced cytidine deaminase"
XX	
PD	WO200058480-A1.
XX	
PD	05-OCT-2000.
XX	
PF	28-MAR-2000; 2000WO-JP01918.
XX	
PR	29-MAR-1999; 99JP-0087192.
PR	24-JUN-1999; 99JP-0178999.
PR	27-DEC-1999; 99JP-0371382.
XX	
PA	(NISB) JAPAN TOBACCO INC.
PI	(HONJ) HONJO T.
PI	Honjo T, Muramatsu M;
XX	
DR	WPI; 2000-611715/58.
DR	P-PSDB; AAB24198.
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as
PT	a target for drug development for immune-related diseases including
PT	allergies -
XX	
PS	Claim 3; Page 135-139; 174pp; Japanese.
XX	
CC	The present sequence encodes human activation-induced cytidine deaminase
CC	(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
CC	has cytidine activity similar to APOBEC-1. AID has antiallergic,
CC	antianemic, antiasthmatic, ophthalmological, anti-HIV and
CC	dermatological activities, and can be used in gene therapy. AID
CC	polynucleotides are useful in methods for identifying drugs for the
CC	treatment of B cell associated immune system disorders, immunodeficiency
CC	diseases and allergies, such as immunoglobulin A (IgA) deficiency
CC	disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
CC	colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
CC	disease, discoid disease, ataxia telangiectasia, common variable
CC	immunodeficiency disorder, MHC (major histocompatibility class
CC	II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC	IgE disorder, and IgG subclass selection disorder. The DNA sequences
CC	encoding AID may be used for gene therapy and the antibodies to the AID
CC	protein may be used for diagnosis and treatment of these disorders.
XX	
SQ	Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;
XX	
Alignment Scores:	
Pred. No.:	4,79e-182 Length: 2818
Score:	198.00 Matches: 198
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	Gaps: 0
US-09-966-880A-8 (1-198) x AAC55312 (1-2818)	
OY	1 Metaspserleueumetafarnargarglyspheleutyrglnphelyasnvalargrtp 20
DB	80 ATGACAGCCTCTTGATGAACCGGAGAGACTTTCCTTACCATTCAAAAATGTCCGGTGG 13
OY	21 Alalysgllyargaragiuhrtyrleucystyrvalvallysarararsperalarhr 40
DB	140 GCTAAGSGTGGCGGTGACACTACCTGTGTCTACGTGTGTAAGAAGCGGTGACAGTGTACA 19
OY	41 Serheserleuasppnegilyrrleuaragnlyasnclcyshstivalgleuleu 60

Dd	200	TCCTTTTCACTGGACCTTGGTATCTCTGGCAATTAAGAACGGCTCCACAGTGGAAATTCTC	259
Qy	61	PhelLeuArgyrrILeserAspTrpAspLeuAspProglyArgCysTyArgValThrTrp	80
Dd	260	TTCTCTCGGCTACATCTCGGACTGGAGCCCTAGACCCTGGCCGCTGCTACCGGGTACCTGG	319
Qy	81	PhethSerTrpSerProCysTyTrpAspCysAlaArgHisValAlaAspPheLeuArgLy	100
Dd	320	TTACCTCTCCGGAGCCCTGCTACGACACTGTGCCGACATGTGGCGCACTTCTGGCGAGGG	379
Qy	101	AspProAsnLeuSerLeuArgTllePheThrAlaArgLeuTyrrPheCysGluAspArgLys	120
Dd	380	AAACCCCAACCTCAGTGTGGAGATCTTCACCGCGCCCTCTCTCTGTGGAGACCGCAAG	439
Qy	121	AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleValIleMetThr	140
Dd	440	GCTAGAGCCCGAGGGGCTGGCGGGCGGTGCACCGCGCGGGGGTGCAAATAGCCATCATGACC	499
Qy	141	PhelAspAspTyrrPheTyrrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys	160
Dd	500	TTCAAGATATATTTTACTGCTGGGAATCTTTTGTAGAAAACCATGAAAGAACTTGTAAA	559
Qy	161	AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgLleLeu	180
Dd	560	GCCTGGGAGAGGCTGCATGAAAATTCAGTCTCTCTCTCCAGACAGCTTCGGCGCATCTCT	619
Qy	181	LeuProLeuTyrrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu	198
Dd	620	TTGGCCCTGTATGAGGTGTATGACTTACAGACAGCATTTGCGACTTGGGACTT	673
RESULT 2			
AAC55317			
ID	AAC55317 standard; DNA: 271 BP.		
XX	AAC55317:		
DT	05-FEB-2001 (first entry)		
XX			
DE	Human activation-induced cytidine deaminase exon 3 seq ID NO:13.		
XX			
KM	Activation-induced cytidine deaminase; AID; cytidine deaminase;		
KM	immune related disease; allergy; allergic disease; antiallergic;		
KM	antichanemic; antichanemic; ophthalmological; anti-HIV; dermatological;		
KM	gene therapy; B cell associated immune system disorder; food allergy;		
KM	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;		
KM	IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;		
KM	drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;		
KM	ataxia telangiectasia; common variable immunodeficiency disorder;		
KM	major histocompatibility class II deficiency disease;		
KM	auto immunodeficiency syndrome; Igs subclass selection disorder; ds.		
XX			
OS	Homo sapiens.		
XX			
FN	W0200058480-A1.		
XX			
PD	05-OCT-2000.		
XX			
PF	28-MAR-2000; 2000WO-JP01918.		
XX			
PR	29-MAR-1999; 99JP-0067192.		
PR	24-JUN-1999; 99JP-0178999.		
XX			
PR	27-DEC-1999; 99JP-0371382.		
XX			
PA	(NISR ) JAPAN TOBACCO INC.		
XX	(HONT/) HONJO T.		
PI	Honjo T, Muramatsu M;		
XX			
DR	WPI; 2000-611715/58.		
XX			
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as		
PT	a target for drug development for immune-related diseases including		
PT	allergies -		

XX PS Claim 18, Page 151, 174pp; Japanese.

CC CC The present invention describes an activation-induced cytidine deaminase (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has cytidine activity similar to APOBEC-1. AID has anti-allergic, CC antianemic, antiasthmatic, ophthalmological, anti-HIV and CC dermatological activities, and can be used in gene therapy. AID CC polynucleotides are useful in methods for identifying drugs for the CC treatment of B cell associated immune system disorders, immunodeficiency CC diseases and allergies, such as immunoglobulin A (IgA) deficiency CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen CC disease, Digeorge disease, ataxia telangiectasia, common variable CC immunodeficiency disorder, MHC (major histocompatibility class) class CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated CC IGE disorder, and Igg subclass selection disorder. The DNA sequences CC encoding AID may be used for gene therapy and the antibodies to the AID CC protein may be used for diagnosis and treatment of these disorders. The CC present sequence represents the exon 3 genomic DNA sequence of human AID.

XX SQ Sequence 271 BP; 47 A; 95 C; 76 G; 53 T; 0 other;

Alignment Scores:

Pred. No.:	2,69e-78	Length:	271
Score:	90.00	Matches:	90
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.45%	Indels:	0
DB:	21	Gaps:	0

US-09-966-880a-8 (1-198) x AAC55317 (1-271)

QY 53 ASNGLYCYSHISVALGULEULEUPHELEUARGTYRILLESASPTRIPASPLEUASPRO 72  
DB 1 AAGGCGTCCACCGTGAATTTGCTCTTCTCCCTACATCTCGGAGCGGACCTAGACCT 60

QY 73 GLIARGCYSTYRARGVALTHRTTPHERTHSERTRPSEPCYSTYRSPCYALARG 92  
DB 61 GCGCGGTGCTACCGCGTCACTGCTGCTGCTGAGCCCGCTGACAGCTGCGCGA 120

QY 93 HISVALAASPHELEUARGLYASNPROASLEUSERLEUARGILEPHERHALARG 112  
DB 121 CATGTGCGCCACTTCTCGAGAGGAGACCCCAACCTCAGTCTGAGATCTTCCACCGCGCC 180

QY 113 LEUTYRPHCYSGIUSPARGLYSALAGIUPROGLUGLYLEUARGRPHLEUHSARGALA 132  
DB 181 CTCTACTCTGTGAGAGCCGCAAGGCTGAGCCCGGAGGCGTGGCGCTGACCGCGCC 240

QY 133 GLIYVALGILNLEAIAILEMETHRPHLEYS 142  
DB 241 GGGGTGCAATATACCATGACCTTCANA 270

RESULT 3  
AAC55314  
ID AAC55314 standard; DNA: 6564 BP.

XX AAC55314:

XX 05-FEB-2001 (first entry)

DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.

XX XX

KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; anti-allergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; Igg subclass selection disorder; ds.

XX OS Homo sapiens.

XX PN WO200058480-A1.

XX PD 05-OCT-2000.

XX PF 28-MAR-2000; 2000WO-JP01918.

XX PR 29-MAR-1999; 99JP-0087192.

XX PR 24-JUN-1999; 99JP-0178999.

XX PR 27-DEC-1999; 99JP-0371382.

XX PA (NIBS) JAPAN TOBACCO INC.

XX PA (HONTU) HONTU T.

XX Hoojo T, Muramatsu M;

XX WPI: 2000-611715/58.

XX Claim 17, Page 145-150, 174pp; Japanese.

PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -

XX SQ Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 other;

Alignment Scores:

Pred. No.:	5.03e-77	Length:	6564
Score:	90.00	Matches:	90
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.45%	Indels:	0
DB:	21	Gaps:	0

US-09-966-880a-8 (1-198) x AAC55314 (1-6564)

QY 53 ASNGLYCYSHISVALGULEULEUPHELEUARGTYRILLESASPTRIPASPLEUASPRO 72  
DB 2592 AAGGCGTCCACCGTGAATTTGCTCTTCTCCGCTACATCTGAGCTGAGACTAGACCT 2651

QY 73 GLIARGCYSTYRARGVALTHRTTPHERTHSERTRPSEPCYSTYRSPCYALARG 92  
DB 2652 GCGCGGTGCTACCGCGTCACTGCTGCTGAGCCCGCTGACAGCTGAGCGCGA 2711

QY 93 HISVALAASPHELEUARGLYASNPROASLEUSERLEUARGILEPHERHALARG 112  
DB 2712 CATGTGCGCCACTTCTCGAGAGGAGACCCCAACCTCAGTCTGAGAGATCTTACCGCGCGC 2771

QY 113 LEUTYRPHCYSGIUSPARGLYSALAGIUPROGLUGLYLEUARGRPHLEUHSARGALA 132  
DB 2772 CTCTACTCTGTGAGAGCCGCAAGGCTGAGCCCGGAGGCGTGGCGCTGACCGCGCC 2851

QY 133 GLIYVALGILNLEAIAILEMETHRPHLEYS 142







KW antihaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
OS Homo sapiens.  
XX W0200058480-A1.  
XX  
XX PD 05-OCT-2000.  
XX  
XX PF 28-MAR-2000; 2000WO-JP01918.  
XX  
XX PR 29-MAR-1999; 99JP-0087192.  
XX PR 24-JUN-1999; 99JP-0178999.  
XX PR 27-DEC-1999; 99JP-0371382.  
XX  
XX PA (NISA ) JAPAN TOBACCO INC.  
XX PA (HONJ/) HONJO T.  
XX  
XX PI Honjo T, Muramatsu M;  
XX  
XX DR WPI; 2000-611715/58.  
XX  
XX PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
XX PS Claim 18; Page 151; 174pp; Japanese.  
XX  
XX CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antihnaemic, antiasthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents the exon 4 genomic DNA sequence of human AID.  
XX  
XX SQ Sequence 116 BP; 32 A; 25 C; 24 G; 35 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 3.13e-28 Length: 116  
Score: 38.00 Matches: 38  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 19.19% Indels: 0  
DB: 21 Gaps: 0  
  
US-09-966-880A-8 (1-198) x AAC55318 (1-116)  
QY 144 TyrPheYrcYsrTrpsnThrPheValGluAsnHisGluArgThrPheIysAlaTrpSiu 163  
DB 3 TATTTTACTGCTGGAACTACTTTGTAGAAACACGAAAGAACTTTCAACCTTGGA 62  
QY 164 G1YleuHisGluAsnSerValArgLeuSerArgGluArgGluLeuLeu 181  
DB 63 GGGCGTGAATAAAATTCAGTCTCTCTCCAGACACTTCGGGCGCATCTTTTG 116  
RESULT 8

AAK81088  
ID AAK81088 standard; DNA; 1665 BP.  
XX  
XX AAK81088;  
AC  
XX  
XX 07-NOV-2001 (first entry)  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35900.  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX OS Homo sapiens.  
XX  
XX PN W0200157182-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 17-JAN-2001; 2001WO-US01354.  
XX  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.  
XX PR 17-MAR-2000; 2000US-0190076.  
XX PR 18-APR-2000; 2000US-0198123.  
XX PR 19-MAY-2000; 2000US-0205515.  
XX PR 07-JUN-2000; 2000US-0209467.  
XX PR 28-JUN-2000; 2000US-0214886.  
XX PR 30-JUN-2000; 2000US-0215135.  
XX PR 07-JUL-2000; 2000US-0216647.  
XX PR 07-JUL-2000; 2000US-0216880.  
XX PR 11-JUL-2000; 2000US-0217487.  
XX PR 11-JUL-2000; 2000US-0217496.  
XX PR 14-JUL-2000; 2000US-0218290.  
XX PR 26-JUL-2000; 2000US-0220963.  
XX PR 26-JUL-2000; 2000US-0220964.  
XX PR 14-AUG-2000; 2000US-0224518.  
XX PR 14-AUG-2000; 2000US-0224519.  
XX PR 14-AUG-2000; 2000US-0225213.  
XX PR 14-AUG-2000; 2000US-0225214.  
XX PR 14-AUG-2000; 2000US-0225265.  
XX PR 14-AUG-2000; 2000US-0225266.  
XX PR 14-AUG-2000; 2000US-0225267.  
XX PR 14-AUG-2000; 2000US-0225268.  
XX PR 14-AUG-2000; 2000US-0225270.  
XX PR 14-AUG-2000; 2000US-0225271.  
XX PR 14-AUG-2000; 2000US-0225447.  
XX PR 14-AUG-2000; 2000US-0225757.  
XX PR 14-AUG-2000; 2000US-0225758.  
XX PR 14-AUG-2000; 2000US-0225759.  
XX PR 18-AUG-2000; 2000US-0226279.  
XX PR 22-AUG-2000; 2000US-0226681.  
XX PR 22-AUG-2000; 2000US-0226686.  
XX PR 22-AUG-2000; 2000US-0227182.  
XX PR 23-AUG-2000; 2000US-0227009.  
XX PR 30-AUG-2000; 2000US-0228924.  
XX PR 01-SEP-2000; 2000US-0229287.  
XX PR 01-SEP-2000; 2000US-0229343.  
XX PR 01-SEP-2000; 2000US-0229344.  
XX PR 01-SEP-2000; 2000US-0229345.  
XX PR 01-SEP-2000; 2000US-0229346.  
XX PR 05-SEP-2000; 2000US-0229309.  
XX PR 05-SEP-2000; 2000US-0229513.  
XX PR 06-SEP-2000; 2000US-0230437.  
XX PR 06-SEP-2000; 2000US-0230438.  
XX PR 08-SEP-2000; 2000US-0231242.  
XX PR 08-SEP-2000; 2000US-0231243.  
XX PR 08-SEP-2000; 2000US-0231244.  
XX PR 08-SEP-2000; 2000US-0231413.  
XX PR 08-SEP-2000; 2000US-0231414.  
XX PR 08-SEP-2000; 2000US-0232080.  
XX PR 08-SEP-2000; 2000US-0232081.  
XX PR 12-SEP-2000; 2000US-0231968.  
XX PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234957.  
PR 25-SEP-2000; 2000US-0234958.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250150.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0234097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SW;  
PI WPI; 2001-483426/52.  
DR  
XX  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX Disclosure; SEQ ID NO 35900; 3071pp + Sequence Listing; English.  
PS  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytosolic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK2169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 1665 BP; 497 A; 343 C; 365 G; 460 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 8.2e-07 Length: 1665  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 8.59% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-966-880A-8 (1-198) x AAK81088 (1-1665)  
  
QY 182 ProLeuTyGluValAspAspLeuArgAspAlaPheArgThrIleuGlyLeu 198  
Db 178 CCCCGTATGAGGTGATGACTTACGACGACGATTCGACTTCGGACTT 228  
  
RESULT 9  
AAK5319  
ID AAK5319 standard; DNA; 2172 BP.  
XX  
XX AAK5319;  
AC  
XX  
DT 05-FEB-2001 (first entry)  
XX  
XX Human activation-induced cytidine deaminase exon 5 SEQ ID NO:15.  
DE  
XX  
XX  
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW



AAC55328  
ID AAC55328 standard; DNA: 30 BP.  
XX  
AC AAC55328;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase PCR primer SEQ ID NO:24.  
XX  
KM Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KM immune related disease; allergy; allergic disease; antiallergic;  
KM antiaesthetic; antisthmatic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease; PCR primer;  
KM auto immunodeficiency syndrome; IgG subclass selection disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000MO-JP01918.  
XX  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
PA (NIBS) JAPAN TOBACCO INC.  
PA (HONJ) HONJO T.  
PI Honjo T, Muramatsu M;  
XX  
DR WPI; 2000-611715/58.  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
XX  
PS Claim 20; Page 158; 174pp; Japanese.  
XX  
CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antinaeemic, antisthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (Iga) deficiency  
CC disease, Iga nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, Digeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC Ige disorder, and Igg subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents a specifically claimed PCR primer for human  
CC AID, which is also used in an example from the present invention.  
XX  
SQ Sequence 30 BP; 7 A; 6 C; 8 G; 9 T; 0 other;

Alignment Scores:  
Pred. No.: 0.125 Length: 30  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.05% Indels: 0  
DB: Gaps: 21

US-09-966-880a-8 (1-198) x AAC55328 (1-30)  
Qy 3 SerLeuLeuMetAsnArgArgLysPheLeu 12  
|||||  
Db 1 AGCCTCTGTGATGACCGAGAGAGATTCTT 30  
RESULT 12  
AAC55323/c  
ID AAC55323 standard; DNA: 27 BP.  
XX  
AC AAC55323;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase PCR primer SEQ ID NO:19.  
XX  
KM Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KM immune related disease; allergy; allergic disease; antiallergic;  
KM antinaeemic; antisthmatic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease; PCR primer;  
KM auto immunodeficiency syndrome; IgG subclass selection disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000MO-JP01918.  
XX  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
PA (NIBS) JAPAN TOBACCO INC.  
PA (HONJ) HONJO T.  
PI Honjo T, Muramatsu M;  
XX  
DR WPI; 2000-611715/58.  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
XX  
PS Claim 20; Page 155; 174pp; Japanese.  
XX  
CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antinaeemic, antisthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (Iga) deficiency  
CC disease, Iga nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, Digeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC Ige disorder, and Igg subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents a specifically claimed PCR primer for human  
CC AID, which is also used in an example from the present invention.  
XX  
SQ Sequence 27 BP; 6 A; 9 C; 7 G; 5 T; 0 other;

## Alignment Scores:

Pred. No.:	1.06	Length:	27
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.55%	Indels:	0
DB:	21	Gaps:	0

US-09-966-880A-8 (1-198) x AAC55332 (1-27)

Qy 93 HisValAlaAspPheLeuArgGlyAsn 101  
 |||  
 DB 27 CATGTGGCCGACCTTCTCGCAGGGAC 1

RESULT 13  
 AAC55332/C  
 ID AAC55332 standard; DNA: 28 BP.  
 XX  
 AC AAC55332;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX

Human activation-induced cytidine deaminase PCR primer SEQ ID NO:28.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease; PCR primer;  
 KW auto immunodeficiency syndrome; Igg subclass selection disorder; ss.

Homo sapiens.

WO200058480-A1.

05-OCT-2000.

28-MAR-2000; 2000WO-JP01918.

29-MAR-1999; 99JP-0087192.

24-JUN-1999; 99JP-0178999.

27-DEC-1999; 99JP-0371382.

(NISB ) JAPAN TOBACCO INC.

(HONO/) HONO T.

Honjo T, Muramatsu M;

WPI; 2000-611715/58.

XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -

Claim 20: Page 160; 174pp: Japanese.

XX The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (Iga) deficiency  
 CC disease, Iga nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class

CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC Ige disorder, and Ige subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders. The  
 CC present sequence represents a specifically claimed PCR primer for human  
 CC AID, which is also used in an example from the present invention.

Sequence 28 BP; 8 A; 6 C; 10 G; 4 T; 0 other;

## Alignment Scores:

Pred. No.:	1.09	Length:	28
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.55%	Indels:	0
DB:	21	Gaps:	0

US-09-966-880A-8 (1-198) x AAC55332 (1-28)

Qy 173 SerArgGlnLeuArgArgIleLeuLeu 181  
 |||  
 DB 27 TCCAGACAGCCTTCGGCGCATCCTTTTG 1

RESULT 14  
 ABLO8467  
 ID ABLO8467 standard; CDNA: 3804 BP.  
 XX  
 AC ABLO8467;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX

Drosophila melanogaster expressed polynucleotide SEQ ID NO 19883.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene: ss.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

P-PSDB; ABB64364.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

Claim 1: SEQ ID NO 19883; 21pp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB157737-AB162072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 3804 BP; 950 A; 1053 C; 1087 G; 714 T; 0 other;

## Alignment Scores:

Pred. No.:	99.4	Length:	3804
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.55%	Indels:	0
DB:	23	Gaps:	0

US-09-966-880A-8 (1-198) x ABL08466 (1-3804)

QY 117 GluAspArgLysAlaGluProGluGly 125  
 |||||  
 DB 709 GAGGACAGAAAGCGCAGCCAGAGGCT 735

## RESULT 15

ABL08466 standard; cDNA; 8797 BP.

AC ABL08466;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 19880.

KW Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB64363.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 genes from Drosophila and for elucidating cell signalling and cell-cell  
 interactions -

PS Claim 1; SEQ ID NO 19880; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 capable of detecting 1000 or more genes from Drosophila. The invention is  
 useful in developmental biology and in elucidating cell signalling and  
 cell-cell interactions in higher eukaryotes for the development of  
 insecticides, therapeutics and pharmaceutical drugs. The invention  
 discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 sequences (AB101840-AB16175) and the encoded proteins  
 (AB57737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 8797 BP; 2327 A; 2150 C; 2113 G; 2207 T; 0 other;

## Alignment Scores:

Pred. No.:	215	Length:	8797
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.55%	Indels:	0
DB:	23	Gaps:	0

US-09-966-880A-8 (1-198) x ABL08466 (1-8797)

QY 117 GluAspArgLysAlaGluProGluGly 125  
 |||||  
 DB 1709 GAGGACAGAAAGCGCAGCCAGAGGCT 1735

Search completed: July 7, 2003, 23:50:07  
 Job time : 225 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 7, 2003, 23:45:38 ; Search time 63 Seconds  
(without alignments)  
963,841 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 198  
Sequence: 1 MDSLMNRKFLYQFNVRW.....ILLPLXVDDLRFDAFTTGL 198

Scoring table:

OLIGO	Xgapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878600

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip  
-O=/cgn2\_1/USPFO.spool/US09966880/runat\_07072003\_142356\_23062/app-query.fasta\_1.391  
-DB=issued\_patents\_NA -QMT=fastap -SUFFIX=olig.rml -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR.SCORE=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTMT=plc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=us09966880.acgn\_1.1\_36 &runat\_07072003\_142356\_23062 -NCPU=6 -ICPU=3  
-NO\_MMAR -LARGEOUTERY -NEG\_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6  
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents.NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	8	4.0	984 1	US-08-470-289-8
2	8	4.0	1006 1	US-08-470-289-5
3	8	4.0	1149 1	US-08-470-289-6
4	8	4.0	1149 1	US-08-470-289-9
5	8	4.0	34303 2	US-08-735-609-4
6	8	4.0	34303 2	US-08-735-609-4
7	8	4.0	34303 3	US-09-315-372-4
8	8	4.0	34303 3	US-09-244-752-4
9	8	4.0	34303 3	US-09-245-497-4
10	8	4.0	34303 3	US-09-562-919-4
11	8	4.0	34382 2	US-08-374-483-6
12	8	4.0	35408 4	US-08-973-334-3

Sequence	Score	Query Match Length	ID	Description
Sequence 3, Appl1	13	8	4.0	35408 4
Sequence 3, Appl1	14	8	4.0	35408 4
Sequence 1, Appl1	15	8	4.0	35935 2
Sequence 1, Appl1	16	8	4.0	35935 2
Sequence 43, Appl1	17	8	4.0	35935 3
Sequence 1, Appl1	18	8	4.0	35935 3
Sequence 1, Appl1	19	8	4.0	35935 3
Sequence 1, Appl1	20	8	4.0	35935 3
Sequence 1, Appl1	21	8	4.0	35935 3
Sequence 1, Appl1	22	8	4.0	35935 3
Sequence 1, Appl1	23	8	4.0	35935 3
Sequence 1, Appl1	24	8	4.0	35935 3
Sequence 1, Appl1	25	8	4.0	35935 3
Sequence 1, Appl1	26	8	4.0	35935 3
Sequence 1, Appl1	27	8	4.0	35935 3
Sequence 1, Appl1	28	8	4.0	35935 3
Sequence 1, Appl1	29	8	4.0	35935 3
Sequence 1, Appl1	30	8	4.0	35935 3
Sequence 1, Appl1	31	8	4.0	35935 3
Sequence 1, Appl1	32	8	4.0	35935 3
Sequence 1, Appl1	33	8	4.0	35935 3
Sequence 1, Appl1	34	8	4.0	35935 3
Sequence 1, Appl1	35	8	4.0	35935 3
Sequence 1, Appl1	36	8	4.0	35935 3
Sequence 1, Appl1	37	8	4.0	35935 3
Sequence 1, Appl1	38	8	4.0	35935 3
Sequence 1, Appl1	39	8	4.0	35935 3
Sequence 1, Appl1	40	8	4.0	35935 3
Sequence 1, Appl1	41	8	4.0	35935 3
Sequence 1, Appl1	42	8	4.0	35935 3
Sequence 1, Appl1	43	8	4.0	35935 3
Sequence 1, Appl1	44	8	4.0	35935 3
Sequence 1, Appl1	45	8	4.0	35935 3

#### ALIGNMENTS

RESULT 1  
US-08-470-289-8  
Sequence 8, Application US/08470299  
Patent No. 5783181  
GENERAL INFORMATION:  
APPLICANT: Murphy, Kay E.  
APPLICANT: Chapman, Conrad G.  
APPLICANT: Clinkenbeard, Helen E.  
APPLICANT: Young, Peter R.  
APPLICANT: Shatzman, Allan R.  
TITLE OF INVENTION: No. 5783181el Compounds  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road, P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,299  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P3100503  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5024  
TELEFAX: 610-270-5090

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 984 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-470-299-8

Alignment Scores:  
Pred. No.: 24.6 Length: 984  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-470-299-8 (1-984)

QY 176 LeuArgArgIleLeuLeuProLeu 183  
Db 837 CTCGACGAGATCCTTCTTCCTCTA 860

RESULT 2  
US-08-470-299-5  
Sequence 5, Application US/08470299  
Patent No. 5783181

GENERAL INFORMATION:  
APPLICANT: Browne, Michael J.  
APPLICANT: Murphy, Kay E.  
APPLICANT: Chapman, Conrad G.  
APPLICANT: Clinkenbeard, Helen E.  
APPLICANT: Young, Peter R.  
APPLICANT: Shatzman, Allan R.  
TITLE OF INVENTION: No. 5783181el Compounds  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road, P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,299  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P31005C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5024  
TELEFAX: 610-270-5090  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1006 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-470-299-5

Alignment Scores:  
Pred. No.: 25.1 Length: 1006  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-470-299-5 (1-1006)

QY 176 LeuArgArgIleLeuLeuProLeu 183  
Db 837 CTCGACGAGATCCTTCTTCCTCTA 860

RESULT 3  
US-08-470-299-6  
Sequence 6, Application US/08470299  
Patent No. 5783181

GENERAL INFORMATION:  
APPLICANT: Browne, Michael J.  
APPLICANT: Murphy, Kay E.  
APPLICANT: Chapman, Conrad G.  
APPLICANT: Clinkenbeard, Helen E.  
APPLICANT: Young, Peter R.  
APPLICANT: Shatzman, Allan R.  
TITLE OF INVENTION: No. 5783181el Compounds  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road, P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,299  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P31005C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5024  
TELEFAX: 610-270-5090  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1149 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-470-299-6

Alignment Scores:  
Pred. No.: 28.6 Length: 1149  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-470-299-6 (1-1149)

QY 176 LeuArgArgIleLeuLeuProLeu 183  
Db 1002 CTCGACGAGATCCTTCTTCCTCTA 1025

RESULT 4  
US-08-470-299-9  
Sequence 9, Application US/08470299  
Patent No. 5783181

GENERAL INFORMATION:  
APPLICANT: Browne, Michael J.  
APPLICANT: Murphy, Kay E.  
APPLICANT: Chapman, Conrad G.



APPLICANT: Clinkenbeard, Helen E.  
APPLICANT: Young, Peter R.  
APPLICANT: Shatzman, Allan R.  
TITLE OF INVENTION: NO. 5783181el Compounds  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporation  
STREET: 709 Swedeland Road, P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,299  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P31005C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5024  
TELEFAX: 610-270-5090  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1149 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-470-299-9

Alignment Scores:  
Pred. No.: 28.6 Length: 1149  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-470-299-9 (1-1149)  
CY 176 LeuAArgIleLeuProLeu 183  
Db 1002 CTCGACGAGTCTTCTCTCTA 1025

RESULT 5  
US-08-735-609-4/c  
Sequence 4, Application US/08735609  
Patent No. 5953560  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,609  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-08-735-609-4

Alignment Scores:  
Pred. No.: 790 Length: 34303  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-735-609-4 (1-34303)  
CY 171 ArgLeuSerArgIleLeuAArg 178  
Db 10380 CGCCTATCAAGACACTCGAGGA 10357

RESULT 6  
US-08-735-609-4/c  
Sequence 4, Application US/08735609  
Patent No. 5954132  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,609  
FILING DATE: 23-Oct-1996  
CLASSIFICATION: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-08-735-609-4

Alignment Scores:  
Pred. No.: 790 Length: 34303  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-735-609-4 (1-34303)

OY 171 ArgleuserArgInleuArg 178

DB 10380 CGCCTATCAAGACACTCAGGAGA 10357

RESULT 7

US-09-315-372-4/c  
Sequence 4, Application US/09315372  
Patent No. 6057158

GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.

APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.

APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco

STATE: California  
COUNTRY: United States of America

ZIP: 94104  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/315,372

FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/735,609

FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"

US-09-315-372-4

Alignment Scores:

Pred. No.: 790

Length:

34303

Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 3 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-315-372-4 (1-34303)

OY 171 ArgleuserArgInleuArg 178

DB 10380 CGCCTATCAAGACACTCAGGAGA 10357

RESULT 8

US-09-244-752-4/c  
Sequence 4, Application US/09244752  
Patent No. 6063622

GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.

APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.

APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco

STATE: California  
COUNTRY: United States of America

ZIP: 94104  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/244,752

FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/735,609

FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"

US-09-244-752-4

Alignment Scores:

Pred. No.: 790 Length: 34303  
Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 4.04% Indels: 0  
DB: 3 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-244-752-4 (1-34303)

OY 171 ArgleuserArgInleuArg 178

DB 10380 CGCCTATCAAGACACTCAGGAGA 10357

RESULT 9  
US-09-245-497-4/c  
Sequence 4, Application US/09245497  
Patent No. 6083750  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/245,497  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/735,609  
APPLICATION NUMBER: <B> FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-09-245-497-4

Alignment Scores:  
Pred. No.: 790 Length: 34303  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
Gaps: 0  
US-09-966-880a-8 (1-198) x US-09-245-497-4 (1-34303)

OY 171 ArgLeuSerArgGlnLeuArgArg 178  
Db 10380 CGCCTATCAAGACACTCAGCAGA 10357

RESULT 10  
US-09-562-919-4/c  
Sequence 4, Application US/09562919  
Patent No. 6451596  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/562,919  
FILING DATE: 02-May-2000  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,609  
FILING DATE: 23-Oct-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-09-562-919-4

Alignment Scores:  
Pred. No.: 790 Length: 34303  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
Gaps: 0  
US-09-966-880a-8 (1-198) x US-09-562-919-4 (1-34303)

OY 171 ArgLeuSerArgGlnLeuArgArg 178  
Db 10380 CGCCTATCAAGACACTCAGCAGA 10357

RESULT 11  
US-08-374-483-6/c  
Sequence 6, Application US/08374483  
Patent No. 5880102  
GENERAL INFORMATION:  
APPLICANT: GEORGE, SAMUEL E.  
APPLICANT: BLAZING, MICHAEL A.  
TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYE P.C.  
STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/374,483  
FILING DATE: 17-JAN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 1579-83  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
TELEX: 200797 NIXN UR  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34382 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-374-483-6

Alignment Scores:  
Pred. No.: 792 Length: 34382  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-374-483-6 (1-34382)

QY 171 ArgLeuserArgGlnLeuArgArg 178  
DB 9800 CGCCTATCAAGACACTCAGAGA 9777

RESULT 12  
US-08-973-334-3/c  
Sequence 3, Application US/08973334  
Patent No. 6261551  
GENERAL INFORMATION:  
APPLICANT: Wilson, James M.  
APPLICANT: Fisher, Krishna J.  
TITLE OF INVENTION: Recombinant Adenovirus and Adeno-  
TITLE OF INVENTION: Associated Virus, Cell Lines, and  
TITLE OF INVENTION: Methods of Production and Use  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Box 457, 321 No. 6261551sttown Road  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release 1.0 Version 1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973,334  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/462,014  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/549,489  
FILING DATE: 27-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNPV012CIPUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 540-9206  
TELEFAX: (215) 540-5818  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35408 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
US-08-973-334-3

Alignment Scores:  
Pred. No.: 815 Length: 35408  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-8 (1-198) x US-08-973-334-3 (1-35408)

QY 171 ArgLeuserArgGlnLeuArgArg 178  
DB 10205 CGCCTATCAAGACACTCAGAGA 10182

RESULT 13  
US-09-563-869A-3/c  
Sequence 3, Application US/09563869A  
Patent No. 6270996  
GENERAL INFORMATION:  
APPLICANT: Wilson, James M.  
APPLICANT: Fisher, Krishna J.  
TITLE OF INVENTION: Recombinant Adenovirus and Adeno-  
TITLE OF INVENTION: Associated Virus, Cell Lines, and  
TITLE OF INVENTION: Methods of Production and Use  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Box 457, 321 No. 62709961sttown Road  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release 1.0 Version 1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/563,869A  
FILING DATE: 03-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/973,334  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/549,489  
FILING DATE: 27-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNPV012CIPUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 540-9206  
TELEFAX: (215) 540-5818  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35408 base pairs  
TYPE: nucleic acid

STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
SEQUENCE DESCRIPTION: SEQ ID NO: 3  
US-09-563-869A-3

## Alignment Scores:

Pred. No.:	815	Length:	35408
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.04%	Indels:	0
DB:	4	Gaps:	0

US-09-966-880A-8 (1-198) x US-09-563-869A-3 (1-35408)

QY 171 ArgLeuSerArgGlnLeuArgArg 178

DB 10205 CGCCTATCAAGACACTCAGAGAGA 10182

## RESULT 14

US-08-549-489-3/c  
Sequence 3, Application US/08549489  
Patent No. 6281010  
GENERAL INFORMATION:  
APPLICANT: Wilson, James M.  
TITLE OF INVENTION: No. 6281010e1 Adenovirus Gene Therapy Vehicle  
TITLE OF INVENTION: and Cell Line  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Box 457, 321 No. 6281010r1etown Road  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/549,489  
CLASSIFICATION: 435  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/462,014  
FILING DATE: 08-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GANPNO13  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 540-9206  
TELEFAX: (215) 540-5818  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35408 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
US-08-549-489-3

## Alignment Scores:

Pred. No.:	815	Length:	35408
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.04%	Indels:	0
DB:	4	Gaps:	0

US-09-966-880A-8 (1-198) x US-08-549-489-3 (1-35408)

QY 171 ArgLeuSerArgGlnLeuArgArg 178

DB 10205 CGCCTATCAAGACACTCAGAGAGA 10182

## RESULT 15

US-08-735-609-1/c  
Sequence 1, Application US/08735609  
Patent No. 5955360  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,609  
CLASSIFICATION: 435  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35935 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-08-735-609-1

## Alignment Scores:

Pred. No.:	827	Length:	35935
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.04%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-8 (1-198) x US-08-735-609-1 (1-35935)

QY 171 ArgLeuSerArgGlnLeuArgArg 178

DB 8877 CGCCTATCAAGACACTCAGAGAGA 8854

Search completed: July 8, 2003, 00:45:48  
Job time : 140 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 7, 2003, 23:46:13; Search time 146 Seconds  
(without alignments)  
2116.961 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 198  
Sequence: 1 MOSLMMNRKFLYQFKNVNM.....ILLPLVEYDDLADARRITGL 198

Scoring table:  
OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1085931 seqs, 780495707 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2169305

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL-frame+ -p2n.model -DEV-x1p  
-Q/cgn2\_1/USPTO.spool/US09966880/runat.07072003.142357.23080/app.query.fasta.1.391  
-DB-published.Applications.NA -OPMT-fastap -SUFFIX-olig.rnpb -MINMATCH=0.1  
-LOOPCT=0 -LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-oligo  
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR-SCORE=quality -THR-MIN=1  
-ALIGN=15 -MODE-LOCAL -OUTFMT-pto -NORM-ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09966880 -CGN\_1.1.122 -runat.07072003.142357.23080  
-NCPU=6 -ICPU=3 -NO-MAP -LARGEQUERY -NEG-SCORES=0 -WAIT -DSPEJOCK=100  
-LONGLOG -DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FCGAPOP=6 -FCGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: Published.Applications.NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PTC\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PTC\_NEW\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	198	100.0	2818	9	US-09-966-880A-7
2	90	45.5	271	9	US-09-966-880A-13
3	90	45.5	6564	9	US-09-966-880A-10
4	90	45.5	11204	9	US-09-966-880A-35

#### ALIGNMENTS

5	49	24.7	148	9	US-09-966-880A-12	Sequence 12, Appl
6	42	21.2	2440	9	US-09-966-880A-1	Sequence 1, Appl
7	38	19.2	116	9	US-09-966-880A-14	Sequence 14, Appl
8	17	8.6	2172	9	US-09-966-880A-15	Sequence 15, Appl
9	10	5.1	30	9	US-09-966-880A-18	Sequence 18, Appl
10	10	5.1	30	9	US-09-966-880A-24	Sequence 24, Appl
11	9	4.5	27	9	US-09-966-880A-19	Sequence 19, Appl
12	9	4.5	28	9	US-09-966-880A-28	Sequence 28, Appl
13	8	4.0	28	9	US-09-966-880A-25	Sequence 25, Appl
14	8	4.0	367	9	US-09-918-995-30146	Sequence 30146, A
15	8	4.0	984	10	US-09-859-361-3	Sequence 3, Appl
16	8	4.0	1056	9	US-10-156-761-7110	Sequence 7110, Ap
17	8	4.0	1188	10	US-09-889-361-1	Sequence 1, Appl
18	8	4.0	1188	10	US-09-859-361-4	Sequence 4, Appl
19	8	4.0	2133	9	US-10-156-761-997	Sequence 997, App
20	8	4.0	13884	10	US-09-070-927A-341	Sequence 341, App
21	8	4.0	29000	9	US-09-906-158-17	Sequence 17, App
22	8	4.0	32480	9	US-09-906-158-23	Sequence 23, Appl
23	8	4.0	34427	9	US-09-847-101B-23	Sequence 2, Appl
24	8	4.0	34408	9	US-10-155-649-3	Sequence 3, Appl
25	8	4.0	35871	10	US-09-956-335-2	Sequence 2, Appl
26	8	4.0	35935	10	US-09-725-720-43	Sequence 43, Appl
27	8	4.0	35935	10	US-09-782-378A-5	Sequence 5, Appl
28	8	4.0	35935	10	US-09-782-378A-5	Sequence 5, Appl
29	8	4.0	35937	10	US-09-782-378A-3	Sequence 3, Appl
30	8	4.0	35978	10	US-09-956-335-1	Sequence 1, Appl
31	8	4.0	36620	9	US-09-953-060-30	Sequence 30, Appl
32	8	4.0	37474	9	US-09-953-060-25	Sequence 25, Appl
33	8	4.0	38519	9	US-09-953-060-28	Sequence 28, Appl
34	8	4.0	81940	9	US-09-755-508B-1	Sequence 1, Appl
35	8	4.0	9025608	9	US-10-156-761-1	Sequence 1, Appl
36	7	3.5	23	9	US-09-964-895-25	Sequence 25, Appl
37	7	3.5	33	9	US-09-941-492-59	Sequence 59, Appl
38	7	3.5	33	9	US-09-756-096A-59	Sequence 59, Appl
39	7	3.5	33	10	US-09-756-095-59	Sequence 59, Appl
40	7	3.5	53	9	US-09-995-225-51	Sequence 51, Appl
41	7	3.5	53	9	US-10-083-168-92	Sequence 92, Appl
42	7	3.5	75	10	US-09-864-761-26823	Sequence 26823, A
43	7	3.5	136	10	US-09-974-300-5651	Sequence 5651, A
44	7	3.5	164	10	US-09-864-761-21666	Sequence 21666, A
45	7	3.5	208	10	US-09-923-876-5781	Sequence 5781, Ap

#### ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
Sequence 7, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CITIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (80)...(673)

```
FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(79)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

Alignment Scores:
Pred. No.: 1.15e-196 Length: 2818
Score: 198.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-7 (1-2818)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTyr 20
Db ATGGACAGCCTCTTGATGAAACCGAGAGAGTTCTTTACCAATCAAAAATGTCGCTGG 139
QY 21 AlAlysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db GCTAAGAGGTGGGCTAGACCTACCTGCTACGTAGTGAAGAGGCGTACAGTGTACA 199
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
Db TCTTTTTCACGTGACCTTTGGTTATCTTCGCAATTAAGAACGCTGCCACGTGAATTGCTC 259
QY 61 PheLeuArgTyrLleSerAspTyrPheLeuAspProGlyArgCysTyrArgValThrTrp 80
Db TTTCTCCGCTACATCTCGAGCTGGGACCTAGACCCCTGGCGCTGCTACCGCTCACCTGG 319
QY 81 PheThrSerTyrSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db TTTACTCTCGGAGAGCCCTGCTAGACCTGCTGCCCGACATGTCGCGACTTTCTGGAGGG 379
QY 101 AsnProAsnLeuSerLeuArgGliePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db AACCCCAACCTCAGTCTGAGAGATCTTCACCGCGGCTCTACTCTGTGAGAGACCGCAAG 439
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnLleAlaLleMetThr 140
Db GGTGACCCCGAGGGGGTGGGGGCTGCACCGCGGGGTCCAAATACCCCTCAATGACC 499
QY 141 PheLeuAspTyrPheTyrCysTyrPheAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db TTTCAAGATATTATTTTACTGCTGGAATACTTTTGTGAAAAACCAATGAAGAATTTTCAA 559
QY 161 AlaTyrGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgLleLeu 180
Db GCTGTGGAAGGGCTGATGAAATTCAGTTCTGCTCTCCAGACAGTGTGGGCGCATCTCT 619
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db TTTGCCCCGTATGAGGTGATGACTTACGAGACGAGCATTTCTGACTTGGGACTT 673

RESULT 2
US-09-966-880A-13
; Sequence 13, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 1999-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
```

```
PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-13

Alignment Scores:
Pred. No.: 5.49e-85 Length: 271
Score: 90.00 Matches: 90
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 45.45% Indels: 0
Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-13 (1-271)

QY 53 AsnGlyCysHisValGluLeuPheLeuArgTyrLleSerAspTyrPheAspPro 72
Db 1 AACGGCTGCCACGTGGAATTTCTTCTCCGCTACATCTGGAGCTGGACCTAGACCT 60
QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTyrSerProCysTyrAspCysAlaArg 92
Db GGGCGCTGCTACCGCGTACCTGTTTACCGCTCGAGACCCCTGCTACGACTGAGCCGA 120
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgGliePheThrAlaArg 112
Db CATGTGGCGCACTTTCTGCGAGGAGACCCCAACCTCAGTCTGAGAGATCTCACCGCGCG 180
QY 113 LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAla 132
Db CTCTACTTCTGTGAGACCGGACGCTGAGCGCGAGGGGCTGGCGGCTGACCGCGGCC 240
QY 181 CTCTACTTCTGTGAGACCGGACGCTGAGCGCGAGGGGCTGGCGGCTGACCGCGGCC 240
QY 133 GlyValGlnLleAlaLleMetThrPheLys 142
Db 241 GGGGTGCAATAGCCATCATGACTTCAA 270

RESULT 3
US-09-966-880A-10
; Sequence 10, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10
; LENGTH: 6564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-10

Alignment Scores:
Pred. No.: 1.04e-83 Length: 6564
Score: 90.00 Matches: 90
Percent Similarity: 100.00% Conservative: 0
```

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 45.45% Indels: 0  
DB: 9 Gaps: 0  
US-09-966-880A-8 (1-198) x US-09-966-880A-10 (1-6564)  
QY 53 AsnGlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro 72  
DB 2592 AACGGCTGCACGTGAATGGCTCTCCCGCTACATCTCGAGACGGAGACCTACACCT 2651  
QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg 92  
DB 2652 GCGCGCTGCACCGCGCTACCTGCTGCTCCTGAGAGCCCGCTGACAGCTGTGCCCA 2711  
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112  
DB 2712 CATGTGGCGACCTTTGTGGAGGGAACCCCAACCTGATCTGAGATCTTCAACCGCGCC 2771  
QY 113 LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgThrPheHisArgAla 132  
DB 2772 CTCTACTTCTGTAGAGACCGCAAGGCTGAGCCCGAGGGGCTGGGCGCTGCACCGCGCC 2831  
QY 133 GlyValGlnIleAlaIleMetThrPheLys 142  
DB 2832 GGGGTGCAAAATAGCATCATGACCTTCANA 2861  
RESULT 4  
US-09-966-880A-35  
; Sequence 35, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasaku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 11204  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880A-35  
Alignment Scores:  
Pred. No.: 1.7e-83 Length: 11204  
Score: 90.00 Matches: 90  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 45.45% Indels: 0  
DB: 9 Gaps: 0  
US-09-966-880A-8 (1-198) x US-09-966-880A-35 (1-11204)  
QY 53 AsnGlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro 72  
DB 7807 AACGGCTGCACGTGAATGGCTCTCCCGCTACATCTCGAGACGGAGACCTACACCT 7866  
QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg 92  
DB 7867 GCGCGCTGCACCGCGCTACCTGCTGCTCCTGAGAGCCCGCTGACAGCTGTGCCCA 7926  
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112  
DB 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112

DB 7927 CATGTGGCGGACCTTCTGCGAGGGAACCCCAACCTGATCTGAGGATCTTCACCGCGGC 7986  
QY 113 LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgThrPheHisArgAla 132  
DB 7987 CTCTACTTCTGTAGAGACCGCAAGGCTGAGCCCGAGGGGCTGGCGGCTGCACCGCGCC 8046  
QY 133 GlyValGlnIleAlaIleMetThrPheLys 142  
DB 8047 GGGGTGCAAAATAGCATCATGACCTTCANA 8076  
RESULT 5  
US-09-966-880A-12  
; Sequence 12, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasaku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 148  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880A-12  
Alignment Scores:  
Pred. No.: 1.78e-42 Length: 148  
Score: 49.00 Matches: 49  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.75% Indels: 0  
DB: 9 Gaps: 0  
US-09-966-880A-8 (1-198) x US-09-966-880A-12 (1-148)  
QY 4 LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGly 23  
DB 2 CTCTTGATGACACCGAGAGAGTCTTACCAATTCAAAATGCGCCGTGAGGTAAGGT 61  
QY 24 ArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSer 43  
DB 62 CGCGGTGAGACCTACCTGCTGCTAGTAGAGAGGGGTGACAGTGTCTACATCTTTCA 121  
QY 44 LeuAspPheGlyTyrLeuArgAsnLys 52  
DB 122 CTGGACCTTGTGTATCTTCCGAATAG 148  
RESULT 6  
US-09-966-880A-1  
; Sequence 1, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasaku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382



```

; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2440
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(686)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(92)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (690)...(2440)
; NAME/KEY: misc_feature
; LOCATION: (1)...(2440)
; OTHER INFORMATION: n = A,T,C or G
US-09-966-880A-1

Alignment Scores:
Pred. No.: 4.7e-34 Length: 2440
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.21% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-1 (1-2440)

QY 54 GTCYAHISVALGLUENUPHELEUARGTYRILLESASPRTSPASPLEUASPROGLY 73
Db 252 GGCCTGCACCTGGAATGTGTGCTCAGCCTACATCTCAGACTGGAGCCTGGACCCGGGC 311
QY 74 ATGCYATYARGLAHTTRTPHETHRSETTPSERPCCYSTYASPCYSALAAHGHIS 93
Db 312 CGGTGTACCGCCTCACCTGCTCCTGAGGCCCTGCTAATGACTGTCCCGGCAC 371
QY 94 VALALA 95
Db 372 GTGGCT 377

RESULT 7
US-09-966-880A-14
; Sequence 14, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 116
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

US-09-966-880A-14
Alignment Scores:
Pred. No.: 4.19e-31 Length: 116
Score: 38.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.19% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-14 (1-116)

QY 144 TYRPHETRYCSTPANTHRPHEVALGLUASNHISGLARGHTRPHELYSALATPGLU 163
Db 3 TATTTTACTGCTGGAATCTTTGTGAAAACCCAGCAAGAACTTCAAGCCTGGAA 62
QY 164 GYLEUHHISGLUASNSERVALARGLEUSERARGSLNLEUARGARGILEAU 181
Db 63 GGCCTGCATGAATAATTCAGTTCTCTCTCCAGACAGCTTGGCGCATCTTTTG 116

RESULT 8
US-09-966-880A-15
; Sequence 15, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 2172
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-15

Alignment Scores:
Pred. No.: 4.94e-08 Length: 2172
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.59% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-15 (1-2172)

QY 182 PROLEUTRYGLUVALASPAPLEUARGASPALAPHEARGHTRLEUAGLYLEU 198
Db 1 CCCCTGATGAGGTGATGACTTACGAGACGCAATTCGTTGGGACTT 51

RESULT 9
US-09-966-880A-18
; Sequence 18, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
```

```

; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially synthesized primer sequence, 22
US-09-966-880A-18

Alignment Scores:
Pred. No.: 0.019      Length: 30
Score: 10.00         Matches: 10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 5.05%      Indels: 0
DB: 9                Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-18 (1-30)

QY 32 ValVallyAaRgArGAsPserAlaThrSer 41
DB 1 GTAGTGAGAGGCGGTGACAGTCTACATCC 30

RESULT 10
US-09-966-880A-24
; Sequence 24, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially synthesized primer sequence, p14
US-09-966-880A-24

Alignment Scores:
Pred. No.: 0.019      Length: 30
Score: 10.00         Matches: 10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 5.05%      Indels: 0
DB: 9                Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-24 (1-30)

QY 3 SerLeuLeuMetAsnArgArGlySpHeleu 12
DB 1 AGCCTCTGTATGAACCGAGAGATTCTT 30
```

```

; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially synthesized primer sequence, 25
US-09-966-880A-19

Alignment Scores:
Pred. No.: 0.19       Length: 27
Score: 9.00          Matches: 9
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 4.55%      Indels: 0
DB: 9                Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-19 (1-27)

QY 93 HisValAlaAspHeleuArgGlyAsn 101
DB 27 CATGGCGCGGACTTCTGCGAGGAGAC 1

RESULT 12
US-09-966-880A-28/c
; Sequence 28, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 28
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially synthesized primer sequence, p26
US-09-966-880A-28
```

Alignment Scores:

Pred. No.:	0.196	Length:	28
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.55%	Indels:	0
DB:	9	Gaps:	0

US-09-966-880A-8 (1-198) x US-09-966-880A-28 (1-28)

QY 173 SerArgGlnLeuArgArgIleLeuLeu 181  
|||||

DB 27 TCCAGACGCTCGCGGCATCTTTG 1

RESULT 13

US-09-966-880A-25/c

; Sequence 25, Application US/09966880A

; Patent No. US20020164743A1

; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku

; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE

; FILE REFERENCE: 06501-088001

; CURRENT APPLICATION NUMBER: US/09/966,880A

; PRIOR FILING DATE: 2001-09-28

; PRIOR APPLICATION NUMBER: PCT/JP00/01918

; PRIOR FILING DATE: 2000-03-28

; PRIOR APPLICATION NUMBER: JP 11-371382

; PRIOR FILING DATE: 1999-12-27

; PRIOR APPLICATION NUMBER: JP 11-178999

; PRIOR FILING DATE: 1999-06-24

; PRIOR APPLICATION NUMBER: JP 11-87192

; PRIOR FILING DATE: 1999-03-29

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 25

; LENGTH: 28

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Artificially synthesized primer sequence, p16

US-09-966-880A-25

Alignment Scores:

Pred. No.:	2.17	Length:	28
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.04%	Indels:	0
DB:	9	Gaps:	0

US-09-966-880A-8 (1-198) x US-09-966-880A-25 (1-28)

QY 44 LeuAspPheGlyTyrLeuArgAsn 51  
|||||

DB 26 CTGAGCTTGTTATCTTCGCAAT 3

RESULT 14

US-09-918-995-30146

; Sequence 30146, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; PRIOR FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/735,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 30146

LENGTH: 367

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-918-995-30146

Alignment Scores:

Pred. No.:	23.2	Length:	367
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.04%	Indels:	0
DB:	9	Gaps:	0

US-09-966-880A-8 (1-198) x US-09-918-995-30146 (1-367)

QY 127 ArgArgLeuHisArgAlaGlyVal 134  
|||||

DB 7 AGGAGCTGCATAGCGCGCGCTG 30

RESULT 15

US-09-859-361-3

; Sequence 3, Application US/09859361

; Patent No. US20020058311A1

; GENERAL INFORMATION:

; APPLICANT: Browne, Michael  
Chapman, Conrad  
Clinkenbeard, Helen  
Robinson, Jeffrey

; TITLE OF INVENTION: Chimeric Leptin Fused to Immunoglobulin  
Domain and Use

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:  
ADDRESS: Smithkline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406

; COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/859,361  
FILING DATE: 17-May-2001  
CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/981,783  
FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P31202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 984 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-859-361-3

Alignment Scores:

Pred. No.:	57.7	Length:	984
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.04%	Indels:	0

DB: 10 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-859-361-3 (1-984)

OY 176 LeuArgArgIleLeuLeuProLeu 183  
|||||  
Db 837 CTCGACGAGATCTTCTCTCTA 860

Search completed: July 8, 2003, 00:47:21  
Job time : 157 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 7, 2003, 23:45:08 ; Search time 1404 Seconds

(without alignments)  
2283.979 Million cell updates/sec

Title: US-09-966-880a-8  
Perfect score: 198  
Sequence: 1 MD5LMMRKKFLYQFKNVM.....ILLPLYVDLDRAPRTGL 198

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xl  
-O/cgnt2.1/USPTO.spool/US09966880/runat.07072003.142356.23048/app.query.fasta.1.391  
-DB=EST -QFMT=fastap -SUFFIX=olig.rst -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US09966880.@CGN.1.1.1525\_@runat.07072003.142356.23048 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WALT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WALT\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gp\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_tam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	198	100.0	743	12	BG686133	BG686133 602638412
2	198	100.0	856	9	AL559877	AL559877 AL559877
3	198	100.0	872	12	BG758510	BG758510 602712721
4	198	100.0	953	14	BQ065440	BQ065440 ABENNCURT
5	198	100.0	1052	14	BQ055935	BQ055935 ABENNCURT
6	193	97.5	820	12	BG757089	BG757089 602715124
7	167	84.3	942	12	BF975166	BF975166 602244657
8	154	77.8	541	12	BF238155	BF238155 601811880
9	142	71.7	889	12	BG686876	BG686876 602650861
10	123	62.1	693	12	BF975096	BF975096 602245679
11	110	55.6	693	12	BG757392	BG757392 602711022
12	42	21.2	522	12	BG144705	BG144705 ut73f07.y
13	42	21.2	653	10	BB637360	BB637360 BB637360
14	30	15.2	442	17	BH302559	BH302559 CH230-100
15	28	14.1	650	9	AT449745	AT449745 AT449745
16	28	14.1	688	9	AT450317	AT450317 AT450317
17	28	14.1	696	9	AT453647	AT453647 AT453647
18	28	14.1	729	9	AT450296	AT450296 AT450296
19	28	14.1	757	9	AT446140	AT446140 AT446140
20	23	11.6	623	9	AT450295	AT450295 AT450295
21	18	9.1	675	10	BB221671	BB221671 BB221671
22	18	9.1	843	12	BG758815	BG758815 602713177
23	18	9.1	928	12	BG398364	BG398364 602439832
24	18	9.1	1034	12	BG755005	BG755005 602711511
25	18	9.1	1140	12	BF664352	BF664352 602146010
26	17	8.6	454	10	AM134750	AM134750 ut-H-B11-
27	14	7.1	939	12	BG341819	BG341819 602463552
28	13	6.6	510	17	A2417616	A2417616 1N0133D22
29	13	6.6	664	12	BG754140	BG754140 602709681
30	10	5.1	670	12	BG341546	BG341546 602463652
31	10	5.1	820	12	BG755526	BG755526 602716206
32	10	5.1	956	12	BF664355	BF664355 602146013
33	9	4.5	202	17	BH121139	BH121139 RRC1-24-3
34	9	4.5	324	9	AV149775	AV149775 AV149775
35	9	4.5	355	12	BG631719	BG631719 CC-estf1cl
36	9	4.5	400	10	AM789252	AM789252 C01013-R
37	9	4.5	431	10	AW566552	AW566552 FK23f11.Y
38	9	4.5	446	12	BF889916	BF889916 289530 MA
39	9	4.5	498	10	BE463445	BE463445 169615 BA
40	9	4.5	538	13	BM036316	BM036316 t178f12.x
41	9	4.5	540	12	BF260401	BF260401 HVSMEF002
42	9	4.5	546	12	BG127344	BG127344 EST472990
43	9	4.5	548	13	BI211882	BI211882 IPL 62 EI
44	9	4.5	626	12	BG465936	BG465936 RH122.46-
45	9	4.5	640	12	BG241527	BG241527 RH122.50-

ALIGNMENTS

RESULT 1  
LOCUS BG686133 743 bp mRNA linear EST 01-MAY-2001  
DEFINITION 602638412F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4766234 5',  
ACCESSION BG686133  
VERSION BG686133.1 GI:13917530  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 743)  
NIH-MGC http://mgc.ncl.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabrs-r@mail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
plate: LCM1626 row: g column: 03  
High quality sequence stop: 740.  
Location/Qualifiers  
1. 743

## FEATURES

## source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4766234"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (Phage-resistant)"  
/note="Organ: B-cells; Vector: pORF7; Site:1. XhoI;  
Site:2. EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

BASE COUNT  
ORIGIN

176 a 197 c 188 g 182 t

## Alignment Scores:

Pred. No.:	2.15e-178	Length:	743
Score:	198.00	Matches:	198
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
	12	Gaps:	0

US-09-966-880A-8 (1-198) x BG686133 (1-743)

QY 1 Metaspserleuauetnsaragatrglyspheleutyrglnphelysasvalargtrp 20  
DB 78 ATGGACAGCCTCTTGATGAAACGGAGAGATTCTTTACCAATTAATAAATGTCGGTGG 137  
QY 21 AAlaysglyaragatrglyuhtyrrleucystyryvalvallysaragarserserlathr 40  
DB 138 GCTAAGGCTGGCGTGAAGCTACTGCTACTGCTAGTAAAGAGCGTGCAGAGTCTACA 197  
QY 41 Serpheserleuaspheglytyrleuargasnlysaanglycyshisvalgluleu 60  
DB 198 TCCCTTTACTGACTTGTGATCTGTCGCAATTAAGACGGGTGCACAGTGAATGCTC 257  
QY 61 Pheleuargtyrllsersasptirpaspheuaspprogllyargcystyryalvalthtrp 80  
DB 258 TTCCTCCGCTACATCTCGAGCTGGAGCTAGACCTGCGCGCTGCTACCGGCTCACCTGG 317  
QY 81 Phehtsertrpserprocystryraspcysalaarghisvalalaaspheleuargly 100  
DB 318 TTCACCTCTCGAGCCCTCGCTACGACTGTGCCCAACATGTGGCCGATTTCTCGAGAGG 377  
QY 101 Aasnproasnleuaserleuarglllephethralaargleutyryphecysgluasparglys 120  
DB 378 AACCCCAACCTAGCTGAGATCTTCAACCGCGCCCTCTACTCTGTGAGAGACCGCAG 437  
QY 121 Alaagluiproglyleuargargleuhsaragalaaglyvalglnllealillemetthr 140  
DB 438 GGTGAGCCCGAGGGCTGGGGGCTGCACCGCGCGGAGTGCATATGCGCATATGAC 497  
QY 141 Pheleysasptirphecystyrrpasnthrphelvalguasnhsigluargtrp 160  
DB 498 TTCAAAGATATTTTACTGCTGGAAATCTTTTGTAGAAACCATGAAAGAACTTTCAA 557

QY

161 AlatrpgluylleuhsigluasnserValArgleuaserarglnleuargargyleu 180  
DB 558 GCGTGGAGAGGCTGCTGTAATAATTCAGTTGCTCTCCAGACAGCTTCGGCCATCCTT 617

QY

181 LeuProleutyrglyuvalaspaspheuargaspalapheargthrlleuglyleu 198  
DB 618 TTGCCCCGTATGAGTGTGATGACTTACGAGACCATTTGTTGGACCT 671

## RESULT 2

AL559877

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

GENOSCOPE

BP 191 91006 EVRY cedex - France

Email: [segreg@genoscope.cns.fr](mailto:segreg@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).

Location/Qualifiers

1. 856

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0D6003YB14"

/clone\_lib="LTI\_FLO11\_BCI"

/sex="male"

/tissue\_type="B-cells from Burkitt lymphoma"

/lab\_host="DH10B"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-stranded cDNA was digested with Not I and cloned  
into the Not I and Eco RV sites of the pCMVSPORT 6 vector.  
Library was constructed by Life Technologies. Contact :  
Feng Liang Life Technologies, a division of Invitrogen  
9800 Medical Center Drive Rockville, Maryland 20850, USA  
Fax : (1) 301 610 8371 Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com>"

BASE COUNT

209 a 217 c 202 g 226 t 2 others

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-966-880A-8 (1-198) x AL559877 (1-856)

QY 1 Metaspserleuauetnsaragatrglyspheleutyrglnphelysasvalargtrp 20  
DB 19 ATGACAGCCTCTTGATGAAACGGAGAGATTCTTTACCAATTAATAAATGTCGGTGG 78  
QY 21 AAlaysglyaragatrglyuhtyrrleucystyryvalvallysaragarserserlathr 40  
DB 79 GCTAAGGCTGGCGTGAAGCTACTGCTACTGCTAGTAAAGAGCGGTGCATGCTACA 138  
QY 41 Serpheserleuaspheglytyrleuargasnlysaanglycyshisvalgluleu 60  
DB 139 TCCCTTTACTGACTTGTGATCTGTCGCAATTAAGACGGGTGCACAGTGAATGCTC 198  
QY 61 Pheleuargtyrllsersasptirpaspheuaspprogllyargcystyryalvalthtrp 80

|||||  
Db 199 TTCCTCCGCTACATCTCGAGCTGGAGACCTGACCCCTGCGCGCTGCTACCCGCTACCTGG 258  
Oy 81 PheThSerTrpSerProCysTyrAspCysAlaArgIstAlaAlaAspPheLeuArgGly 100  
Db 259 TTCACCTCCGAGACCCCTGCTGAGCTGCGCGAGATGCGCGACTTTCGCGAGGG 318  
Oy 101 AsnProAsnLeuSerLeuArgIlePheThAlaArgLeuTyrPheCysGluAspArgLys 120  
Db 319 AACCCCAACCTCATCTGAGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCCGAG 378  
Oy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyAlaGlnIleAlaIleMetThr 140  
Db 379 GCTGAGCCCGAGGGCTGCGCGGCTGACCGCGCGCGGTGCAAAATAGCAATCAGAGAC 438  
Oy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
Db 439 TTCAAAATATTATTTTACTGCTGGAATCTTTGTAAACCAATGAAAGAACTTCAAA 498  
Oy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
Db 499 GCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCCTT 558  
Oy 181 LeuProLeuTyrGluValAlaAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
Db 559 TTGCCCTGTATGAGTGTGATGACTTACGAGACGCAATTGCTACTTGGGACTT 612  
RESULT 3  
Bg758510 872 bp mRNA linear EST 15-MAY-2001  
LOCUS 602712721P1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4853069 5'  
DEFINITION mRNA sequence.  
ACCESSION Bg758510  
VERSION Bg758510.1 GI:14069163  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 872)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1061698 row: 1 column: 06  
High quality sequence stop: 836.  
Location/Qualifiers  
1. 872  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4853069"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

BASE COUNT  
ORIGIN

211 a 221 c 212 g 228 t

Alignment Scores:  
Pred. No.: 2 6e-178 Length: 872  
Score: 198.00 Matches: 198  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 12 Gaps: 0  
US-09-966-880a-8 (1-198) x Bg758510 (1-872)

Oy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
Db 66 ATGAGACGCTCTTGATGACCGAGAGAGATTCTTTACCAATTCAAAAATGTCGGCG 125  
Oy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
Db 126 GCTAAGGGTCGGCGTGAAGCTACCTGCTGCTACGTAAGTGAAGAGCGTGAAGTGTACA 185  
Oy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
Db 186 TCCTTTCTACTGAGCTTGTGTTATCTTCGCAATGAAGAGCGGTGCAGATTTGCTC 245  
Oy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
Db 246 TTCCTCCGCTACATCTCGAGCTGGAGCTAGACCTGGCCGCTGCTACCGGCTACCTGG 305  
Oy 81 PheThSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
Db 306 TTCACCTCCTGGAGCCCGCTGCTAGACATGTGCCGCAACATGTGGCGACTTCTGCGAGG 365  
Oy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
Db 366 AACCCCAACCTCAGTGTGAGATCTTCACCGCGCGCTCTACTTCTGAGGACCGCAAG 425  
Oy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyAlaGlnIleAlaIleMetThr 140  
Db 426 GCTGAGCCCGAGGGCTGGCGGCTGCACCCCGCGGGGTGCAAAATAGCACTCTGAGAC 485  
Oy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
Db 486 TTCAAAATATTATTTTACTGCTGAGATCTTTGTGAAACCAATGAAAGAACTTCAAA 545  
Oy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
Db 546 GCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCCTT 605  
Oy 181 LeuProLeuTyrGluValAlaAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
Db 606 TTGCCCTGTATGAGTGTGATGACTTACGAGACGCAATTCGACTTGGGACTT 659  
RESULT 4  
BQ065440 953 bp mRNA linear EST 02-APR-2002  
LOCUS BQ065440  
DEFINITION AGENCOURT\_6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:592977  
5', mRNA sequence.  
ACCESSION BQ065440  
VERSION BQ065440.1 GI:19894486  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 953)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Lou Staudt  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation





Db 244 TTCCCTCGCTACATCTCGGACTGGACCTAGACCCCTGCGCTCTACTCCGCTCACCCTGG 303  
 QY 81 PheThSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 Db 304 TTCACCTCTCTGAGCCCTGCTAGACTGCTCCCGACATGTGGCGGACTTCTGCGAGGG 363  
 QY 101 AsnProAsnLeuSerLeuArgGlyPheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 Db 364 AACCCCAACCTCAGTCTAGATCTTACCGCGCGCTCCTACTCTGTAGAGACCGGAG 423  
 QY 121 AlaGluProGluGlyLeuArgGlyLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 Db 424 GCTGAGCCCGAGGGGCTGGCGGCTGCACCGCCCGGGGCAAAATAGCATCATGACC 483  
 QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 Db 484 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCAAGAAAGAACTTTGAAA 543  
 QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 Db 544 GCCTGGGAAGGGCGCATGAATAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGCATCTT 603  
 QY 181 LeuProLeuTyrGluValAlaAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
 Db 604 TTGCCCCGTATGAGGTGATGACTTACGAGACCATTTGCTACTTGGGACTT 657

RESULT 6  
 LOCUS Bg757089 820 bp mRNA linear EST 15-MAY-2001  
 DEFINITION 602715124F1 NIH\_MGC\_48 Homo sapiens CDNA clone IMAGE:485517 5',  
 mRNA sequence.  
 ACCESSION Bg757089  
 VERSION Bg757089.1 GI:14067742  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 820)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L10M1704 row: O column: 06  
 High quality sequence stop: 675.  
 Location/Qualifiers  
 1..820  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:485517"  
 /clone\_1db="NIH\_MGC\_48"  
 /tissue\_type="Primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pGB7; Site: 1: XhoI;  
 Site: 2: EcoRI; CDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GCGACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-CDNA Synthesis Kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

BASE COUNT 210 a 230 c 200 g 180 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	141e-173	Length:	820
Score:	193.00	Matches:	193
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.47%	Indels:	0
DB:	12	Gaps:	0

US-09-966-880a-8 (1-198) x Bg757089 (1-820)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 Db 85 ATGACAGCCTCTGTATGAACCGAGAGAGTTCTTTTACCAATTCAAAAATGTCGCGG 144  
 QY 21 ALVLYGLYARGARGGLUHRTRYRLEUCYSTYRVALVALLYARGARGASPERALATHR 40  
 Db 145 GCTAAGGCTGGGGTGAAGCTACCTGCTCTAGTAGTAAAGAGCGTGAAGTCTTACA 204  
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 Db 205 TCTTTTTCACGTGACTTTTGTTATCTTCCCAATAGAACGGCTGCGACGTGAATTCCTC 264  
 QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 Db 265 TTCCCTCGCTACATCTCGGACTGGAGCTAGACCTGAGCCGCTGCTACCGCTCACCCTGG 324  
 QY 81 PheThSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 Db 325 TTCACCTCTCTGAGCCCTGCTAGACTGCTGCCCGACATGTGGCGGACTTCTGCGAGGG 384  
 QY 101 AsnProAsnLeuSerLeuArgGlyPheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 Db 385 AACCCCAACCTCAGTCTGAGATCTTCACCGCGCGCTCTACTTGTGAGGACCGCAAG 444  
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 Db 445 GCTGAGCCCGAGGGGCTGGCGGCTGCACCGCCCGGGGTCCAAATAGCATCATGACC 504  
 QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 Db 505 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCAAGAAAGAACTTTGAAA 564  
 QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 Db 565 GCCTGGGAAGGGCGCATGAATAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGCATCTT 624  
 QY 181 LeuProLeuTyrGluValAlaAspAspLeuArgAspAlaPhe 193  
 Db 625 TTGCCCCGTATGAGGTGATGACTTACGAGACCATTTG 663

RESULT 7  
 LOCUS Bf975166 942 bp mRNA linear EST 22-JAN-2001  
 DEFINITION 602244657F1 NIH\_MGC\_48 Homo sapiens CDNA clone IMAGE:4335639 5',  
 mRNA sequence.  
 ACCESSION Bf975166  
 VERSION Bf975166.1 GI:12342381  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 942)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.lnl.gov>  
 Plate: LCM1207 row: a column: 16  
 High quality sequence stop: 707.  
 Location/Qualifiers  
 source  
 1..942

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4335639"  
 /clone\_lib="NIH\_MGC\_48"  
 /issue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library."

BASE COUNT 249 a 209 c 251 g 233 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	9,92e-149	Length:	942
Score:	167.00	Matches:	180
Percent Similarity:	99.45%	Conservative:	0
Best Local Similarity:	99.45%	Mismatches:	0
Query Match:	84.34%	Indels:	1
DB:	12	Gaps:	0

US-09-966-880A-8 (1-198) x BF975166 (1-942)

```

QY 1 MetasperlleuMetasnaArgLysPheleuTygln-PheLysasnValArgtr 20
DB 78 ATGGACAGCCTCTGATGAACCGAGGAACTTCTTACCAAGTCAAAATGTCGGCTG 137
QY 20 PALATSGLYARARGLUTHTYRLEUCYSTYRVALVALLYARGARGSPERIALAH 40
DB 138 GGCCTAAGGTCGGCTAGACCTTACCTGCTACGRTAGAAAGCGCTACAGTCTCTAC 197
QY 40 rSerPheSerleuAspPheGlyTYRleuArgasnLYsasnGLYcysHISvalGluLeu 60
DB 198 ATCCTTTCACTGACCTTGTATCTTCCATATAGAAAGCTGCCACGTGSAATTCCT 257
QY 60 uPheLeuArgTYRleSerAspTRPaspLeuAspProGlyArgCYSTYRArgValThr 80
DB 258 CTTCCTCCGCTACATCTCGGACCTGCGACCTAGACCTGGCGCTGCTACCGCTCAC 317
QY 80 pPheThrSerTRPserProCYSTYRAspCYsAlaArgHISvalAlaAspPheLeuArg 100
DB 318 GTTCACCTCTCTGGAGCCCTGCTGACAGCTGTGCCGACAGTGGCGACTTCTGCGAG 377
QY 100 yAsnProAsnLeuSerleuArgTlePheThrAlaArgLeuTYRPhcCYsGluAspArg 120
DB 378 GAACCCCAACCTCAGTGTGAGATCTTACACCGCGCGCTTACTTCTGTGAGGACCGCA 437
QY 120 sAlaGlnPProGlnGlyLeuArgArgLeuHISArgAlaGlyValGlnIleAlaIle 140
DB 438 GCGTGAGCCCGAGGGGCTCGGCGGCTGCACGACGCGGGGTCCAATATGACATGAC 497
QY 140 rPheLeuAspTYRPhcTYRTrpAsnThrPheValGluAsnHISGlnArgThrPhe 160
DB 498 CTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCAAGAAAGAACTTCAA 557
QY 160 sAlaTrpGlnGlyLeuHISGluAsnSerValArgLeuSerArgGlnLeuArgArgLeu 180
DB 558 AGCTGGGAAGGCTGCATGAATTCAGTTCTCTCTCCAGACACACTTGGCGCATCTCT 617
QY 180 u 180
DB 618 T 618

```

RESULT 8  
 BP238155  
 LOCUS  
 DEFINITION  
 accession  
 version  
 keywords  
 source  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 541)  
 NIH-MGC <http://mgc.ncl.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bsh-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.lnl.gov>  
 Plate: LCM895 row: p column: 20  
 High quality sequence stop: 541.  
 Location/Qualifiers  
 source  
 1..541

## FEATURES

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4054915"  
 /clone\_lib="NIH\_MGC\_48"  
 /issue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library."

## BASE COUNT

120 a 150 c 142 g 129 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,26e-136 Length: 541  
 Score: 154.00 Matches: 154  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 77.78% Indels: 0  
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BF238155 (1-541)

```

QY 1 MetasperlleuMetasnaArgLysPheleuTyglnPheLysasnValArgtr 20
DB 78 ATGGACAGCCTCTGATGAACCGAGGAACTTCTTACCAATTCAAAATGTCGGCTG 137
QY 21 AlaLysGlyArgArgGlnThrTYRleuCYSTYRValValLYsArgARGAspSerAlaThr 40
DB 138 GCTAAGAGGTGGCGTGAACCTACCTGTGCTACAGTAGAAGAGCGGAGAGGTACACA 197
QY 41 SerPheSerleuAspPheGlyTYRleuArgasnLYsasnGLYcysHISvalGluLeu 60
DB 198 TCCCTTTTACAGGACTTGTGTTATCTTCCATATAGAAAGCGCTGCCAGTGGATTCCTC 257
QY 61 PheLeuArgTYRleSerAspTRPaspLeuAspProGlyArgCYSTYRArgValThr 80
DB 258 TTCTTCGCTACATCTCGGACTGGAGACTAGACCTGCGCTGCTACAGCGGTACCTGG 317

```

QY 81 PheThSerTPSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 DB 318 TTCACTCTCTGAGCCCTGCTAGACAGCTGCCGACATGTGGCCGACTTTCTCGGAGG 377  
 QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 DB 378 AACCCCACTCAGTCTGATGATCTTCACCCGCGCTCTACTCTGTGAGACCGCAG 437  
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValAlaGlnIleAlaIleMetThr 140  
 DB 438 GCTGAGCCCGAGGGGCTGCGGCTGACCGCGCGGGGCAATAGCATCATGACC 497  
 QY 141 PheLysAspTyrPheTyrCysTyrPasnThrPheValGluAsn 154  
 DB 438 TTCAAGATTTATTTTACTGCTGGAATCTTTGTAGAAAC 539  
 RESULT 9  
 BG686876 889 bp mRNA linear EST 01-MAY-2001  
 LOCUS 602650861F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763247 5'  
 DEFINITION mRNA sequence.  
 ACCESSION BG686876  
 VERSION BG686876.1 GI:13918273  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 889)  
 NIH-MGC http://mgc.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LNCMI618 row: j column: 16  
 High quality sequence start: 6  
 High quality sequence stop: 727.  
 Location/Qualifiers  
 1. 889  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="4763247"  
 /clone\_id="NIH\_MGC\_48"  
 /tissue\_type="Primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pORF7; Site: 1; XhoI;  
 Site-2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GCCACGAG(9). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 205 a 223 c 257 g 203 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.17e-125 Length: 889  
 Score: 142.00 Matches: 142  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 71.72% Indels: 0  
 Ds: 12 Gaps: 0

US-09-966-880a-8 (1-198) x BG686876 (1-889)  
 QY 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTyr 20  
 DB 38 ATGGACAGCTCTTGTATGAAACCGAGAGAGTTCTTTACCAATGCAAAATATCCGCTGG 97  
 QY 21 AlaLysGlyArgArgGlyLutThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
 DB 98 GCTAAGGTCGCGGTAGACCTTCTGCTGCTACGTAGTAAAGGCGGTACGTCTTACA 157  
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 DB 158 TCCTTTCACTGAGACTTGTGTTATCTTCCCAATAGAAAGCGGTGCACGTGAATGCTC 217  
 QY 61 PheLeuArgTyrIleSerAspTyrPasnThrPheValGluAsn 80  
 DB 218 TTCTCTCGGTACATCTCGGACTGCGACTAGACCCCTGGCGCTGCTACCGCTCACCTGG 277  
 QY 81 PheThSerTPSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 DB 278 TTCACTCTCTGAGCCCTGCTAGACAGCTGTGCCGACATGTGCCGACTTCTGCGAGG 337  
 QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 DB 338 AACCCCACTCAGTCTGATGATCTTCACCCGCGCTTACTTCTGTGAGACCGCAG 397  
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 DB 398 GCTGAGCCCGAGGGGCTGCGGCTGACCGCGCGGGGTCAATATGACATCATGACC 457  
 QY 141 PheLys 142  
 DB 458 TTCAAA 463  
 RESULT 10  
 BF975096 693 bp mRNA linear EST 22-JAN-2001  
 LOCUS 602245679F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:436722 5'  
 DEFINITION mRNA sequence.  
 ACCESSION BF975096  
 VERSION BF975096.1 GI:12342311  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 693)  
 NIH-MGC http://mgc.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LNCMI205 row: n column: 19  
 High quality sequence stop: 692.  
 Location/Qualifiers  
 1. 693  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="436722"  
 /clone\_id="NIH\_MGC\_48"  
 /tissue\_type="Primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pORF7; Site: 1; XhoI;  
 Site-2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GCCACGAG(9). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC library."

BASE COUNT 179 a 167 c 162 g 185 t  
ORIGIN

## Alignment Scores:

Pred. No.:	5,86e-107	Length:	693
Score:	123.00	Matches:	123
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	62.12%	Indels:	0
DB:	12	Gaps:	0

US-09-966-880a-8 (1-198) x Bf975096 (1-693)

QY 76 TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAla 95  
Db 2 TACCGGCTCACCTGGTTCACCTCTGAGCCCTGACAGCTGTCACAGCTGCCGACATGTGGCC 61  
QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115  
Db 62 GACTTCTGCGAGGAGAACCCCAACCTCAGCTGAGATCTTCACCCGGCCCTTACTTC 121  
QY 116 CysGluAspArgTyrAlaGluProGluGlyLeuArgLeuHisArgAlaGlyValGln 135  
Db 122 TGTGAGAGCCGCAAGGCTGAGCCCGAGGGGCTCGCGGCTGACCCCGCGGGGTGCNA 181  
QY 136 IleAlaIleMetThrPheLeuAspTyrPheTyrCysTrpAsnThrPheValGluAsnHis 155  
Db 182 ATAGCCATCATGACCTCAAGATATTTTACTGCTGGAATCTTTGTAGAAACCAT 241  
QY 156 GluArgThrPheLeuSalatrrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175  
Db 242 GAAGAAGCTTTCAAGGCTGGGAGGCGCTGCATGAATAATTCAGTTGCTCTCCAGACAG 301  
QY 176 LeuArgArgIleLeuLeuProLeuTyrGluValAspAspLeuArgSpAlaPheArgThr 195  
Db 302 CTTCGGCGCATCTCTTTGGCCCTGTATGAGTTGATGACTTACGAGACCATTTGCTACT 361  
QY 196 LeuGlyLeu 198  
Db 362 TTGGGACTT 370

RESULT 11  
Bg757392 693 bp mRNA linear EST 15-MAY-2001  
LOCUS 602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851580 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg757392  
VERSION Bg757392.1 GI:14068045  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 693)  
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabos-remail.nih.gov](mailto:cgabos-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM1694 row: k column: 05  
High quality sequence stop: 693.

FEATURES  
source

Location/Qualifiers  
1..693  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4851580"  
/clone\_11b="NIH\_MGC\_48"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pORF7; Site:1: XhoI;  
Site:2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC library."

BASE COUNT 157 a 188 c 178 g 170 t  
ORIGIN

## Alignment Scores:

Pred. No.:	1.43e-94	Length:	693
Score:	110.00	Matches:	161
Percent Similarity:	98.77%	Conservative:	0
Best Local Similarity:	98.77%	Mismatches:	1
Query Match:	55.56%	Indels:	2
DB:	12	Gaps:	0

US-09-966-880a-8 (1-198) x Bg757392 (1-693)

QY 35 ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGly 54  
Db 184 AGCGTGACAGTGTCTACATCTTTTCTACGTGACTTTGGTTATCTTCGCAATAAAGAGCC 243  
QY 55 CysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArg 74  
Db 244 TGCCACGTGGAAATGCTCTTCTCCCGTACATCTCGGACTGGAGCTGGACCTAGACCCCGGCCG 303  
QY 75 CysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisVal 94  
Db 304 TGCTACCCCGTACCTGTGCTCCTCGGAGCCCTGCTACGACTGTGCGCGGACATGTG 363  
QY 95 AlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyr 114  
Db 364 GCGGACTTGTGCGAGGAGACCCCAACCTCAGTGTGAGATCTTCACCGCGCGCTTAC 423  
QY 115 PheCysGluAspArgTyrAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyVal 134  
Db 424 TTCTGTGAGGACCGGAAGCGTACCGAGGGGCTGGCGCGCTGCACCGCGCGGGGTG 483  
QY 135 GluIleAlaIleMetThrPheLeuAspTyrPhe-TyrCysTrpAsnThrPheValGluAs 154  
Db 484 CAATAGCCATCAGACCTTCCTCAAGATTATCT-TTACTGTGCGAATCTTTGTAGAAAA 542  
QY 154 nHisGluArgThrPheLeuSalatrrpGluGlyLeuHisGluAsnSerValArgLeuSerArg 174  
Db 543 CCATGAAGAAGACTTTCAGAGCTGGGAGGCGCTGCATGAATAATTCAGTTGCTCTCCAG 602  
QY 174 GlnLeuArgArgIleLeuLeuProLeuTyrGluValAspAspLeuArgSpAlaPheArg 194  
Db 603 ACAGCTTGGCGGACCTCTTTGGCCCTGTATGAGGTGATGACTTACGAGACGCAATTTCG 662  
QY 194 gThrLeu 196  
Db 663 TACTTTG 669  
RESULT 12  
Bg144705 522 bp mRNA linear EST 01-FEB-2001  
LOCUS Bg144705  
DEFINITION ut73f07.v1 Soares\_mouse\_NMG2\_bcell Mus musculus cDNA clone  
IMAGE:333637 5' similar to TR:09WV00 09WV00 ACTIVATION-INDUCED  
CYTIDINE DEAMINASE.;, mRNA sequence.  
ACCESSION Bg144705

VERSION	ESF.	BGI44705.1	GI:12648105
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	1 (bases 1 to 522)		
TITLE	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> This clone is available royalty-free through LNL; contact the IMAGE Consortium ( <a href="mailto:info@image.lnl.gov">info@image.lnl.gov</a> ) for further information. MGI:1077801 Seq primer: -40RP from Gldbo High quality sequence stop: 487.		
FEATURES			
Source	Location/Qualifiers		
	1..522		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:3333637"		
	/clone_1fb="Soares mouse, NMGB, bcell"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: germinal B-cell; Vector: pT73D-Pec (pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCACATCTGAGTGGAGCGCGCGCTGTTTATTTTATTTTATTTTATTTT T 3'] : double-stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized; constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	113 a 140 c 140 g 129 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	6.47e-30	Length:	522
Score:	42.00	Matches:	42
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	21.21%	Indels:	0
DB:	12	Gaps:	0
US-09-966-880A-8 (1-198) x BGI44705 (1-522)			
QY	54	GLGYSHSVAlGULenLeuPheLeuArgYrILleSerASPTrPASPLeuAsPProGly	73
Db	77	GGCGCCACGGGAGATTGTGTCTCTACGACACCTCTGAGCTGGAGCTGGACCGCGGC	136
QY	74	ARGCYSTRARGVALThTrPrPheThSerTrpSerPProCYSTYASPCysAlaArgHis	93
Db	137	CGGtGtTACCCCGtCACCTGStTtACtCCtCGAGCCGtGtATGATGtGCCCGGCAC	196
QY	94	ValAla 95	
Db	197	GTGGCT 202	
RESULT 13			
LOCUS	BB637360		
DEFINITION	BB637360 RIKEN full-length enriched, adult male aorta and vein Mus	653 bp	mRNA linear EST 26-OCT-2001
ACCESSION	BB637360		
VERSION	BB637360.1		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 653)		
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		

AUTHORS	<p>Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanaagaki,T., Hara,A. Himoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasak ,D., Shibata,K., Shingawa,A., Shiraki,T., Sobue,Y., Suzuki,H., Tatematsu,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.</p>
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL	Unpublished (2001)
COMMENT	<p>Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayasui,N., Sugahara,Y., Shibata,K., Itch ,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagi,K., Fujikake,S., Inoue,K., Togawa,Y., Itawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamataka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (<a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a>) for further details. e mouse tissues.</p>
FEATURES	Location/Qualifiers
SOURCE	<pre> 1 . 653    /organism="Mus musculus"    /dp_xref="taxon:10090"    /clone_id="A530070C03"    /clone_1db="RIKEN full-length enriched, adult male aorta and vein"    /sex="male"    /tissue_type="aorta and vein"    /dev_stage="adult"    /lab_host="DH10B"    /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGAGATCCACAGACTCTTTTTTTTTTTTTVN 3'], cDNA was prepared by using triethose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGATTTCGAGTTAAATAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I."</pre>
BASE COUNT	155 a 198 c 155 g 145 t
ORIGIN	



Mon Jul 14 14:47:13 2003

us-09-966-880a-8.Orig.rst

---

Page 11